Departamento de Sistemas y Automática

TRABAJO FIN DE GRADO

Expresión emocional mediante sonidos no verbales en robots sociales

Autor: Miguel Álvarez Guerrero

Director: Javi Fernánedz de Gorostiza Luengo
Tutor: María de los Ángeles Malfaz Vázquez

Madrid, Junio 2016

Título: Expresión emocional mediante sonidos no verbales en robots sociales
Autor: Miguel Álvarez Guerrero
Director: Javier Fernández de Gorostiza
Tutor: María de los Ángeles Malfaz Vázquez

EL TRIBUNAL

Presidente:

Vocal:

Secretario:

Realizado el acto de defensa y lectura del Trabajo Fin de Grado el día ..... de ..... de ... en ....., en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de:

VOCAL

SECRETARIO PRESIDENTE

Agradecimientos

Agradezco a Frank Grimes "Graimito" su ejemplo de lucha frente a las advesidades.

Resumen

Este trabajo se enmarca dentro del estudio de la interacción humano-robot, en la cual el robot y el ser humano se relacionan teniendo en cuenta las normas sociales y un contexto de interacción. En esta interacción las emociones juegan un papel fundamental, permitiendo una comunicación natural y efectiva y aportando al robot una apariencia más cercana al usuario. Por otro lado, el uso de expresiones orales no verbales permite un mayor grado de universalidad en la interacción. Nuestros objetivos son tres. El primero comprende la creación de una aplicación software que genere respuestas emocionales orales no verbales, a partir de la interacción entre humano y robot. El segundo objetivo se basa en el desarrollo de expresiones emocionales sonoras mediante síntesis de sonido. El tercer objetivo trata de implementar las tecnologías desarrolladas en un robot llamado Maggie, para probar su funcionamiento.

Palabras clave: Interacción humano-robot, expresión oral no verbal, emociones primarias, síntesis de sonidos.

Abstract

This work is a part of the study of human-robot interaction (HRI), in which the robot and the human being relate taking into account social norms and interaction context. In this interaction, emotions play a key role, allowing a natural and effective communication and providing to the robot a closer look to the user. Furthermore, the use of non-lingüistic utterances (NLUs) allows a greater degree of universality in the interaction. Our goals are three. The first involves the creation of a software application that generates emotional responses through NLUs from the interaction between human and robot. The second goal is based on the development of emotional utterances through sound synthesis. The third goal is to implement the technologies developed in a robot named Maggie, to test its working.

Keywords: HRI, NLU, basic emotions, sound synthesis.

Contenidos

Lista de Figuras

Lista de Tablas

Capítulo 1

Introducción

Un robot social es aquel que interactúa y se comunica con las personas de forma natural y efectiva, siguiendo comportamientos, patrones y normas sociales [2]. La interacción humano-robot (en adelante HRI, del inglés *Human-Robot Interaction*) es posible siempre que los niveles de comunicación estén adaptados al contexto en el que ésta se produzca.

Estos niveles de comunicación son dos: un nivel explícito o verbal, en el que tiene lugar la transmisión de la información verbal propia del idioma; y un nivel implícito o no verbal, donde la comunicación no se realiza con palabras, sino por medio del lenguaje corporal, la gestualización, la expresión facial y la expresión oral [3]. Nuestro trabajo se centrará en este segundo nivel comunicativo, concretamente en la expresión oral.

Para realizar este trabajo, nos hemos apoyado en la perspectiva evolutiva que ofreció Charles Darwin en su libro *"La expresión de las emociones en el hombre y en los animales"*, según la cual la expresión oral no verbal (en adelante NLU, del inglés *Non Lingüistic Utterance*) está íntimamente relacionada con la expresión de las emociones [4]. Nótese que la estrecha vinculación entre apariencia corporal, facial o vocal y estado interno llevó a Darwin a utilizar el término *expresión*, término que desde entonces ha sido utilizado hasta nuestro días.

Darwin enfatizó la función expresiva de las emociones en los animales, que no requerían de un lenguaje para poder comunicar sus estados internos a sus semejantes, sino que por medio de la expresión emocional podían obtener ayuda de ellos. Por tanto, las NLUs podría considerarse como un *lenguaje primitivo* y universal para los miembros de una determinada especie. En la especie humana, actualmente, esta función de ayuda a la supervivencia parece vestigial, habiéndose transformado, en el devenir de nuestra historia, en una herramienta de comprensión interindividual y formación de relaciones sociales.

En una investigación actual llevada a cabo por el psicólogo Alber Mehrabian, profesor emérito en UCLA, en la que se estudiaba la importancia comunicativa entre los distintos niveles de comunicación a la hora de expresar emociones [3], se encontró que en ciertas situaciones en las que la comunicación verbal era altamente ambigua, el 7% de la información era atribuible a las palabras, el 38% a la voz, y el 55% al lenguaje corporal.

Volviendo a la HRI, podemos entender ahora que la comunicación afectiva a través de las NLUs cobra una gran importancia. Si las relaciones personales entre individuos se forman a través de la comunicación de sus estados de ánimo, de sus deseos e intenciones, la interacción entre el hombre y el robot debe nutrirse de este estilo de interacción, y la efectividad de ésta depende, como decíamos al principio, de una correcta adaptación al

contexto social en que se produzca.

En el presente trabajo hemos querido enfocar el estudio de la HRI en la comunicación oral no verbal, debido a las ventajas que aporta. Entre estas ventajas podemos destacar que la comunicación no verbal es, como decíamos antes, universal y reconocible por todos los individuos de una especie, lo cual permite entre ellos una interacción directa, sencilla y efectiva.

1.1 Motivación del proyecto

Este trabajo forma parte de un gran proyecto cuyo principal objetivo es la construcción de un robot social autónomo. Al ser social, una de las características requeridas del robot es que su modo de interacción sea parecida a la de los humanos con los que se comunica [5]. Debido a esto, la sensación al interactuar con el robot será de que está "vivo", de que no es una simple máquina. No hay que perder de vista que la esencia de la robótica social es que la naturaleza social no es una funcionalidad más del robot, sino una de sus características esenciales.

Por tanto este trabajo pretende, a través de la expresión no verbal de estados afectivos, favorecer la interacción entre el hombre y el robot. Nuestra tarea ha consistido en crear una simulación de expresión emocional durante la interacción del usuario con Maggie, un simpático robot que cuenta con una interfaz táctil a través de la cual se puede jugar a juegos.

Gracias a la aportación del Departamento de Automática y Robótica de la Universidad Carlos III de Madrid, que ha ofrecido la posibilidad de realizar el proyecto en Maggie, y a Javi F Gorostiza, por permitir el uso del juego digital "Kill the Roaches" creado por él, hemos podido implementar el entorno de interacción a partir del cual desarrollar nuestro trabajo.

Veamos cuáles son los objetivos de este proyecto:

- Desarrollar una aplicación software orientada a la interacción Humano-Robot, que capte estímulos del entorno, los procese y transforme en patrones de expresión sonora no verbal en tiempo real. Concretamente, se probará la utilidad de esta aplicación usando como ambiente estimular un juego digital en formato Flash, que será arrancado en un Tablet PC.
- Generar mediante síntesis de sonido un grupo de NLUs que expresen emociones, a través de la tecnología Operator del *software* Ableton Live 9 Suite. Estas NLUs serán emitidas durante el juego, en tiempo real, en función del rendimiento del jugador.
- Implementar el sistema de interacción (juego + aplicación software + sonidos generados por síntesis) en Maggie, un robot social desarrollado por el RoboticsLab del Departamento de Ingeniería de Sistemas y Automática de la Universidad Carlos III de Madrid.

1.2 Estado del arte

Antes de adentrarnos en el desarrollo de nuestro proyecto, hemos hecho un análisis del estado del arte en lo referente a las NLUs en robots sociales, y sus repercusiones en la

interacción Humano-Robot. El contenido mostrado en este apartado está dividido en tres partes:

- En la primera describimos brevemente las características de los robots sociales más representativos que utilizan la expresión oral para comunicar emociones, tanto a nivel implícito como explícito.
- En la segunda parte mostramos los resultados de dos estudios sobre expresión oral emocional en humanos, los cuales nos han servido para analizar las características sónicas en la expresión de emociones básicas.
- En la tercera parte comentamos los resultados de estudios con robots sociales, en los que se ha investigado los efectos de la expresión oral implícita de las emociones en la HRI.

1.2.1 Expresión oral emocional en robots sociales

La robótica social es una reciente rama de la robótica, desarrollada a partir de la década de los 90 por investigadores de robótica e Inteligencia Artificial. Llegada esta nueva visión de la robótica, el papel de los robots alcanza un nivel superior: el social, donde las emociones juegan un papel esencial en el reconocimiento y en la interacción con humanos [6]. La HRI abarca actualmente un amplio campo de desarrollo y es fuente de numerosas investigaciones, que van desde el reconocimiento facial y vocal hasta la expresión verbal y no verbal. En concreto, en el desarrollo de robots mascota puede verse el papel fundamental de las NLUs. A continuación veremos unos cuantos ejemplos.

PARO

PARO es un robot interactivo desarrollado por AIST, una de las empresas japonesas líderes en automatización industrial [7]. La apariencia adorable de PARO (ver Figura 1.1) y su gama de comportamientos ofrece los beneficios de la terapia con animales a distintos colectivos de personas (mayores, con discapacidades, con síntomas de demencia y niños hospitalizados). Se ha encontrado que la interacción con PARO reduce el estrés de los pacientes y sus cuidadores, aumenta su motivación y estimula la interacción entre éstos, promoviendo la socialización. Es, según el Guinness World Records, el robot más terapéutico del mundo.

Figura 1.1: PARO

PARO puede aprender cómo comportarse del modo en el que el usuario prefiera, y responder a su nuevo nombre. En lo referente a la comunicación oral, puede imitar la voz de una cría de foca real.

Aibo

Aibo [8] representa toda una gama de robots mascota desarrollados por SONY, desde su primer modelo lanzado en el año 2000, el Aibo ERS-110, pasando por el modelo ERS-210

con apariencia felina, hasta su versión más sofisticada, el Aibo ERS-7, de 2005 (ver Figura 1.2).

Desde los modelos primera y segunda generación de AIBO se pueden cargar diferentes paquetes de software comercializados por Sony. AIBOware es el título dado al software AIBO. El AIBOware life permite que el robot comience su interacción como cachorro, y se vaya desarrollando hasta alcanzar el estado de adulto. En este desarrollo influirá la interacción que su propietario tenga con él. El AIBOware Explorer permite la interacción con un robot completamente maduro capaz de entender 100 comandos de voz.

La comunicación emocional de Aibo con humanos se basa en las NLUs, por medio de *earcons* [1] con los que puede simular placer, alegría, sorpresa y otros estados afectivos.

(a) Aibo ERS-110 (b) Aibo ERS-210 (c) Aibo ERS-7

Figura 1.2: Aibo

NeCoRo

NeCoRo [9] es otro robot terapéutico con apariencia de gato (Figura 1.3). Fue desarrollado por Omron, una compañía electrónica japonesa, en el año 2004, y responde a las necesidades terapéuticas y de interacción con su amplia gama de conductas expresivas de tipo gestual, facial y vocal.

NeCoRo tiene un mecanismo de generación de expresiones emocionales que le permite mostrar satisfacción, enfado, deseos de dormir o ser acariciado, de acuerdo con los ritmos fisiológicos. Puede generar 48 vocalizaciones de gato distintas y, mediante mecanismos de aprendizaje, adaptar sus conductas expresivas de independencia o necesidad de atención (su personalidad) en función del feedback que le proporcione su dueño.

Figura 1.3: NeCoRo

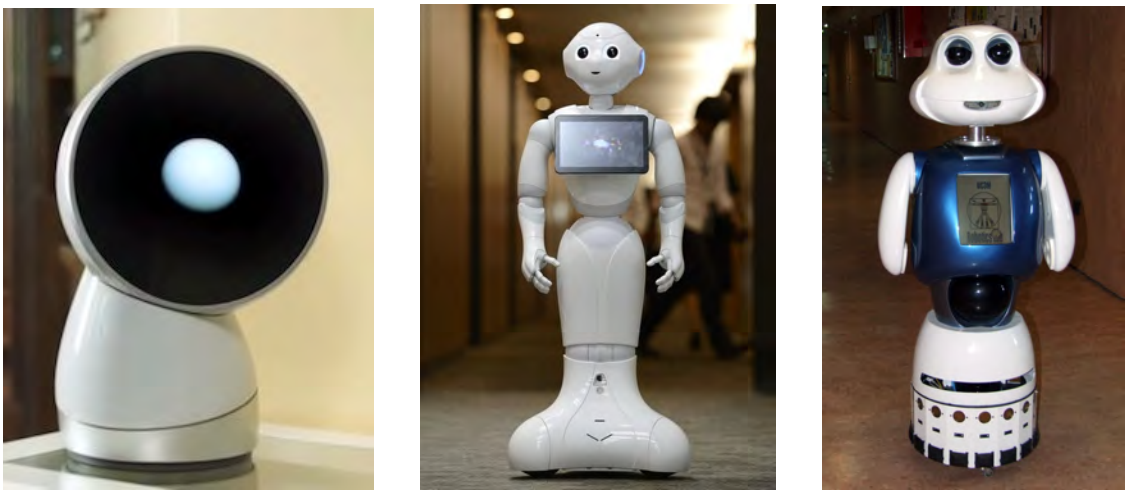
[1] Un *earcon* es un sonido distintivo y breve que representa un evento específico o transmite otra información. El término viene del inglés, y es una transposición de la palabra *icon* (icono), de la modalidad visual a la auditiva.

Robots con expresión verbal

No pasaremos al siguiente punto sin comentar, aunque sea brevemente, los nuevos avances en expresión emocional verbal, representados en robots como Jibo, Pepper o Maggie (Figura 1.4).

Jibo [10] es un robot pensado para la familia, con un sistema de comunicación no verbal que le permite adaptarse naturalmente a la interacción con el usuario, y a responder con claves expresivas emocionales que se ajustan al contexto y de ese modo ser uno más de la familia.

Pepper [11] es considerado el robot humanoide que mejor implementa la comunicación emocional, adaptando su comportamiento al estado de ánimo de su interlocutor. Para comunicarse, Pepper se ayuda del lenguaje, pero sobre todo utiliza la comunicación no verbal de los gestos o la expresión oral y la información audiovisual del tablet PC del pecho para mostrar emociones. Pepper evoluciona con el ususario, memoriza gradualmente los tratos personales con el individuo y se adapta a sus hábitos.



(a) Jibo (b) Pepper (c) Maggie

Figura 1.4: Robots con expresión verbal

Maggie, el robot del RoboticsLab de la Universidad Carlos III de Madrid, ha sido utilizado en numerosas investigaciones en relación a la HRI. En Maggie cofluye el trabajo de profesores y doctorandos de la universidad, pero no sólo de ellos, sino también aportaciones de los alumnos del máster en robótica, y de estudiantes de primer y segundo ciclo. Más adelante, en el capítulo 3, en el que explicamos las plataformas y tecnologías utilizadas, nos detendremos a explicar sus características y el uso que le hemos dado en el proyecto.

1.2.2 Características sónicas de la expresión de emociones básicas

En este apartado comentamos dos estudios que nos han servido como análisis de las características sónicas en la expresión emocional:

- El primero es uno de los estudios transculturales más exhaustivos realizados hasta la fecha, llevado a cabo por Scherer y Wallbott en 1994 [12], y en el que se estudiaron 7 emociones básicas en 37 países de todos los continentes. En este estudio, 2921 participantes informaron sobre sus propios comportamientos expresivos realizados ante cada una de las 7 emociones.

- En el segundo estudio, realizado por Scherer, Johnstone y Klasmeyer en el año 2003 [13], se evalúan los efectos de cada emoción sobre diferentes parámetros vocales, como la fluencia, la frecuencia principal, la prosodia, el esfuerzo vocal y el tipo de fonación.

A pesar de que estos estudios se realizaron sobre 7 emociones básicas, nosotros sólo mostramos los resultados de aquellas emociones que hemos incluido en nuestro modelo, que son: *alegría*, *asco*, *miedo* y *tristeza*, aparte de un estado afectivo de base, la *calma*, que explicaremos más adelante.

Los resultados de estos dos estudios han sido reflejados en las Tablas 1.1 a 1.5:

Parámetros acústicos	**Efecto**
Fluencia	
Nº de sílabas por segundo	$\geq$[a]
Duración de la sílaba	$\leq$
Duración de las vocales acentuadas	$\geq$
Nº y duración de las pausas	$<$
Frecuencia principal (F0)[b] ***y Prosodia***	
F0 media	$>$
F0 desviación	$>$
F0 rango	$>$
Frecuencia de sílabas acentuadas	$\geq$
Gradiente de ascenso y descenso de F0	$>$
Esfuerzo vocal y tipo de fonación	
Intensidad (dB) media	$\geq$
Intensidad (dB) desviación	$>$
Gradiente de ascenso y descenso de la intensidad	$\geq$

Tabla 1.1: Efecto de la alegría en los parámetros acústicos.

[a]Nota: en fonemas específicos, < "menor", "más bajo", "más lento"; = "igual" o "neutral"; > "mayor", "más alto", "más rápido"; <> ambos "menor" y "mayor" se han obtenido en los estudios.

[b]F0 hace referencia a la frecuencia con la que vibran las cuerdas vocales.

Comportamientos expresivos acústicos	Porcentaje
Conducta verbal	
Reír	4,2
Llorar/sollozar	6,6
Chillar/gritar	7,4
Otros cambios de voz	16,7
Conducta paralingüística	
Cambios en la melodía	10,7
Interrupciones en el habla	7,2
Cambios en la velocidad del habla	10,8
Conducta no verbal	
Silencio	38,8
Verbalizaciones cortas	21,7
Verbalizaciones largas	14,2

Tabla 1.2: Porcentajes de comportamientos expresivos realizados ante la emoción de asco.

Parámetros acústicos	Efecto
Fluencia	
Nº de sílabas por segundo	>
Duración de la sílaba	<
Duración de las vocales acentuadas	<
Nº y duración de las pausas	<>
Frecuencia principal (F0) y Prosodia	
F0 media	>
F0 desviación	>
F0 rango	>
Gradiente de ascenso y descenso de F0	<>

Tabla 1.3: Efecto del miedo en los parámetros acústicos.

Comportamientos expresivos acústicos	Porcentaje
Conducta verbal	
Reír	4,3
Llorar/sollozar	15,5
Chillar/gritar	12.8
Otros cambios de voz	18,4
Conducta paralingüística	
Cambios en la melodía	11,6
Interrupciones en el habla	15,2
Cambios en la velocidad del habla	12,8
Conducta no verbal	
Silencio	50,5
Verbalizaciones cortas	20,1
Verbalizaciones largas	7,3

Tabla 1.4: Porcentajes de comportamientos expresivos realizados ante la emoción de miedo.

Parámetros acústicos	Efecto
Fluencia	
Nº de sílabas por segundo	$<$
Duración de la sílaba	$>$
Duración de las vocales acentuadas	$\geq$
Nº y duración de las pausas	$>$
Frecuencia principal (F0) y Prosodia	
F0 media	$<$
F0 desviación	$<$
F0 rango	$<$
Frecuencia de sílabas acentuadas	$<$
Gradiente de ascenso y descenso de F0	$<$
Esfuerzo vocal y tipo de fonación	
Intensidad (dB) media	$\leq$
Intensidad (dB) desviación	$<$
Gradiente de ascenso y descenso de la intensidad	$<$

Tabla 1.5: Efecto de la tristeza en los parámetros acústicos.

1.2.3 Interacción emocional humano-robot

En este apartado comentaremos los resultados de tres estudios en los que se han estudiado NLUs en robots.

En el primer estudio, realizado por Robin Read y Tony Belpaeme sobre la expresión emocional a partir de expresiones no verbales, y en el cual utilizaban a niños en la interacción con el robot [14], se sacaron las siguientes conclusiones:

- La interpretación por parte del niño de la expresión emocional emitida por el robot

está bastante influida por el contexto y otras modalidades de expresión, como los gestos y las expresiones faciales del propio robot.

- La clasificación emocional que hace el niño del estado interno del robot a partir de sus expresiones emocionales suele responder a categorías cerradas y exclusivas, es decir, el robot está o bien triste, o bien contento, o asustado, o en calma. Estos resultados apoyan el enfoque categorial dado a las emociones por buena parte de los investigadores de la emoción.
- El tono en la expresión no verbal suele ser entendido por el niño como un predictor del grado de activación emocional o arousal [2] del robot, y no tanto de la categoría emocional en la que éste se halle.

En el estudio que realizó Robin Read para realizar su tesis doctoral, en el cual investigó sobre NLUs en la HRI [15], se obtuvieron las siguientes conclusiones:

- Cuando se les presenta a distintos participantes la misma expresión no verbal, aislada de cuaquier contexto, cada una es susceptible de ser asignada a una emoción distinta.
- Participantes de distintas edades suelen interpretar la expresión emocional dentro de una categoría cerrada, al igual que veíamos en los niños de los anteriores estudios.
- El contexto situacional sesga en gran medida la interpretación que hará el participante del significado emocional de la expresión no verbal emitida por el robot.

Por último, en una investigación llevada a cabo por el Trinity College sobre el papel de la prosodia y entonación para expresar emociones [16], las conclusiones fueron similares.

1.3 Estructura del documento

Para facilitar la lectura del documento, se resume a continuación el contenido de cada capítulo:

- En el ***Capítulo 2*** se realiza una explicación de las disciplinas y enfoques de investigación que hemos tenido en cuenta a la hora de realizar este trabajo. Este capítulo se divide en dos grandes bloques: la Psicología de la Emoción, en el que se hace referencia a las NLUs y a los enfoques desde los que se han estudiado las emociones; y la síntesis de sonidos, en el que se hace un breve recorrido por las distintas técnicas de síntesis y conceptos relacionados.
- En el ***Capítulo 3*** se enumeran y explican cada una de las plataformas y tecnologías que hemos utilizado, pretendiendo justificar debidamente por qué las hemos escogido y el uso que les hemos dado.
- En el ***Capítulo 4*** se hace referencia al sistema propuesto, haciendo una descripción estructural y funcional, por un lado, y explicando la implementación de las tecnologías desarrolladas en las plataformas, por otro.
- En el ***Capítulo 5*** se exponen las conclusiones del trabajo, teniendo en cuenta los objetivos tomados, y evaluando si éstos ha sido cumplidos.

[2] A lo largo de este documento los términos *arousal* y *activación emocional* serán tratados como sinónimos, si bien la literatura científica sobre el estudio de las emociones encuentra diferencias entre ellos.

Capítulo 2

Base Teórica

2.1 Psicología de la Emoción

Como comentábamos en la introducción, la comunicación afectiva a través de la expresión oral es necesaria para que la interacción entre el humano y el robot sea natural y efectiva.

La comunicación emocional se encuadra dentro de la Psicología de la Emoción, el área de la Psicología que estudia la emoción como conducta, englobando en ella los cambios fisiológicos asociados [17], la experiencia subjetiva (o sentimiento) del individuo [18], y el afrontamiento desplegado por éste ante la situación que ha incitado tal emoción [19]. Además, la Psicología de la Emoción, a través del enfoque del Procesamiento de la Información, entiende la emoción como un sistema de análisis y procesamiento de inputs tanto interoceptivos como exteroceptivos [1], vinculándola de modo directo a la dimensión cognitiva del sujeto [20].

Antes de entrar a explicar los aspectos expresivos de las emociones, nos gustaría clarificar brevemente el propio concepto de emoción, por cuanto ha sido muchas veces fuente de ambigüedades y malentendidos. Para ello haremos al lector la siguiente pregunta.

2.1.1 ¿Percibimos las emociones o sólo sus efectos?

Los datos que utilizamos para explicar la presencia de vivencias emocionales son de tipo subjetivo-fenomenológico (el sentimiento de alegría o enfado), fisiológico (la activación corporal) y expresivo-motor (los gestos faciales, la prosodia...). Pero tal y como exponen acertadamente teóricos como Peter Lang [21] o Arne Öhman [1], ninguno de estos datos es, por sí solo, señal incuestionable de que la emoción está siendo experimentada.

Nadie ha visto nunca la tristeza, o la alegría. Sólo observamos a alguien que ríe, llora o grita, o que nos dice que se siente alegre o enfadado, o medimos su actividad cardíaca o el nivel de activación del hemisferio izquierdo o del derecho. Las emociones no se pueden observar directamente, pues sólo se infieren a partir de los datos de emoción (en la Figura 2.1 puede verse la relación entre estos datos y la emoción).

En lo referente a nuestro proyecto -teniendo en cuenta que en él sólo nos centramos en datos emocionales expresivos-, si bien la expresión vocal emitida por el robot puede crearnos la idea de que está expresando con ella una emoción, esta idea será cierta en la

[1] Un estímulo o input *interoceptivo* es aquél que sucede dentro del medio interno del organismo; un estímulo *exteroceptivo* sucede en el medio externo, es decir, en el ambiente.

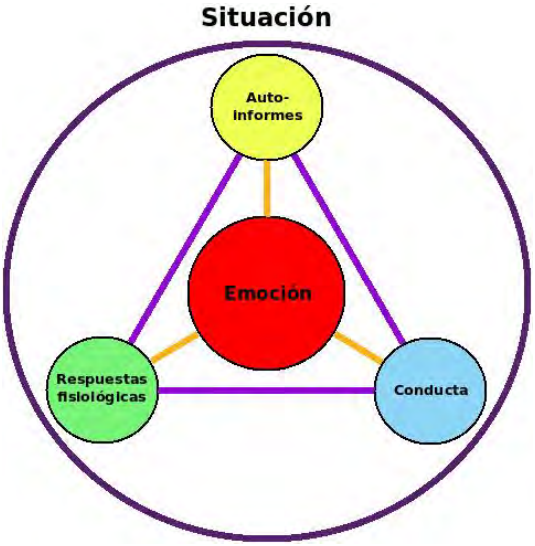


Figura 2.1: Los estados emocionales se infieren a partir de eventos observables empíricamente. Los trazos color naranja señalan esa característica inferencial de la emoción. Las líneas color morado destacan las relaciones entre eventos observables, los datos de emoción (Adaptado de Öhman [1]).

medida en que entendamos que el robot interactúa con nosotros en base a una progamación previa, es decir, *que está programado para responder así*, pero nunca podemos suponer que el propio robot *experimenta* de forma subjetiva tal emoción. No obstante, permítasenos denominar en adelante a este tipo de expresiones como expresiones emocionales, dado que que expresan, si bien de manera sintética o artificial, reacciones a situaciones concretas que en interacción entre humanos generarían respuestas emocionales.

Tras esta aclaración sobre lo que entendemos por emoción, exponemos a continuación y de forma breve los aspectos expresivos de las emociones.

2.1.2 Expresión emocional oral no verbal

Para poder desarrollar nuestras NLUs por medio de la síntesis de sonido, es importante conocer antes el fundamento teórico que da cuenta de los elementos o parámetros sonoros que definen estas expresiones. Las disciplinas que estudian la comunicación no verbal son la Paralingüística, la Kinesia, la Proxémica, la Cronémica y el Comportamiento Diacrítico (en la Figura 2.2 sólo aparecen las más estudiadas).

Este trabajo se centra en la Paralingüística -ya que nuestro estudio se limita a la comunicación oral no verbal de la emoción-, y para abordarla hemos tomado como referencia la visión que aporta Fernando Poyatos [22]. Este linguista define el paralenguaje como el conjunto de vocalizaciones y cualidades de la voz, dentro del cual se incluyen todos aquellos sonidos que no se ajustan a la estructura fonética de la lengua. Veamos a continuación los elementos que componen estos sonidos.

Elementos paralingüísticos

En un nivel elemental dentro del marco de la expresión paralingüística, nos encontramos con las *cualidades primarias*, que son las características físicas del sonido que individualizan a una persona. Según Poyatos [23], estas cualidades están condicionadas por diferentes factores:

- Biológicos: sexo y edad.

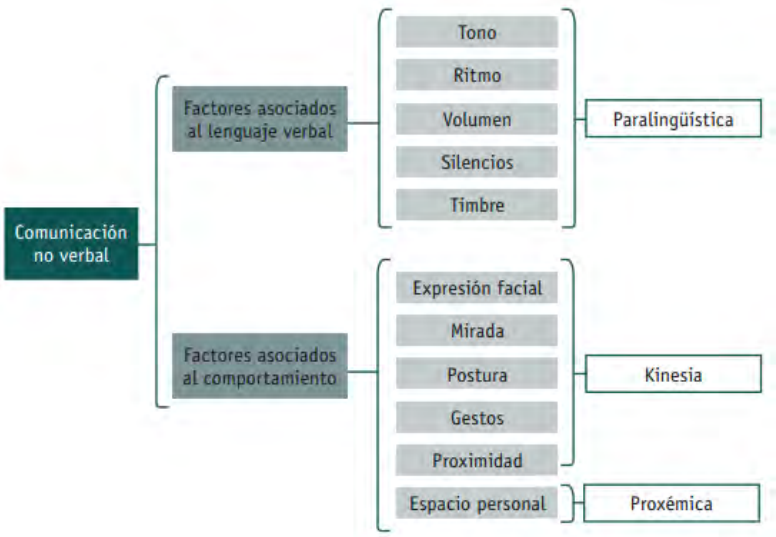


Figura 2.2: Factores asociados al lenguaje verbal y al comportamiento

- Fisiológicos.
- Psicológicos: personalidad.
- Socioculturales.
- Ocupacionales.

Entre las cualidades primarias que destaca este autor, las que más han aportado a nuestro estudio han sido:

1. *Tono e inflexión*: son dos factores íntimamente relacionados que relacionan la emoción y la expresión que empleamos, sirviendo como reguladores de éstas dos. El tono y la inflexión que se emplean al hablar permiten comunicar actitudes, intenciones y emociones del hablante. El tono se entiende como el atributo psicológico del sonido que lo caracteriza como agudo o grave (alto o bajo), en función de la propiedad física *frecuencia*. En cuanto a la inflexión, distinguimos tres tipos, que podrán variar en función del contexto cultural en el que sea ejecutada la expresión:
   (a) Ascendente: expresa duda, indecisión, inseguridad. También puede asociarse al miedo.
   (b) Descendente: transmite firmeza, determinación y confianza.
   (c) Mixto: sugiere ironía y sarcasmo.
2. *Ritmo*: en términos paralingüísticos, es la fluidez verbal con la que se expresa una persona. También puede entenderse como un análogo del ritmo que compone una pieza musical. La *cadencia* sucede cuando el ritmo es regular.
3. *Volumen*: se relaciona con la intensidad con la que expresamos. Lo empleamos para poner énfasis, regular e incluso alterar un proceso de comunicación. Generalmente, un volumen bajo nos indicará tristeza, calma o aburrimiento. Por el contrario, un volumen alto transmite alegría, miedo o excitación. El volumen está íntimamente relacionado con el nivel de arousal fisiológico.

4. *Timbre*: es el registro que nos permite distinguir una fuente de sonido de forma inmediata. El timbre depende de la cantidad de armónicos que componen el sonido de la voz, y de la intensidad de cada uno de ellos. Un timbre abierto es aquel en el que predominan armónicos con frecuencias altas, y un timbre cerrado, en el que predominan las frecuencias bajas. En la vocalización humana, la ejecución de las vocales en sentido abierto-cerrado sería: A, E, I, O, U. También se hace distinción entre timbre áspero y redondeado.

Una vez explicados los elementos que modulan la expresión emocional, veamos a continuación un punto que ha sido clave para nuestro trabajo: las distintas aproximaciones al estudio de las emociones que se han llevado a cabo en la historia de la Psicología de la Emoción.

2.1.3 Enfoques al estudio de las emociones

Si bien en un principio el *enfoque dimensional* y el *enfoque categorial* en el estudio de las emociones parecían dos modos distintos y contrapuestos de ver la configuración emocional, con el paso de los años y tras la gran acumulación de estudios se ha optado por verlos como complementarios. Parece que cada enfoque ofrece un marco de investigación concreto que puede llegar hasta distintos puntos en la descripción y explicación de elementos y mecanismos de las emociones y que, en conjunto, ambas explicaciones ofrecen una visión globalizadora y más completa que cada una por separado.

Veamos a continuación la aportación de cada enfoque a la comprensión de las emociones.

El enfoque dimensional

Según la teoría de Schachter y Singer [24], también conocida como "del arousal más cognición", la activación fisiológica es condición necesaria para que se produzca una emoción, pero no suficiente. Esta activación será inespecífica en sí misma en ausencia del otro elemento necesario, el evaluativo-cognitivo, el cual determinará la cualidad de la emoción en base a la interpretación que de el sujeto a la situación (creencias, etiquetas verbales o indicadores de contexto).

Las investigaciones recogidas bajo esta visión *dimensional* de las emociones describen y localizan las emociones en un espacio continuo delimitado por tres ejes:

- La *valencia afectiva* que va del placer al displacer, y que permite diferenciar las emociones en función de su tono hedónico sea positivo o negativo.
- La *activación o arousal* que va de la calma al entusiasmo, y que permite diferenciar las emociones por la intensidad de los cambios fisiológicos entre las condiciones de tranquilidad o relajación, y el de extrema activación o pánico incontrolable.
- El *control* que va del extremo controlador *de* la situación, al extremo controlado *por* la situación, y que permite diferenciar las emociones en función de quién ejerza el dominio, la persona o la situación desencadenante.

La Figura 2.3 muestra el mapa emocional que más investigación ha generado hasta el momento [25], en el que se tienen en cuenta sólo las dimensiones de *valencia afectiva*

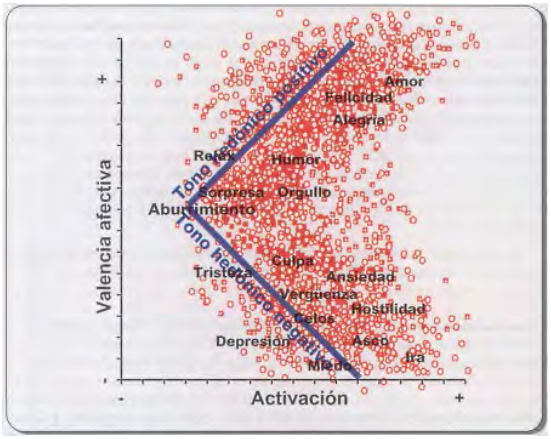


Figura 2.3: Espacio emocional bidimensional

y *arousal*. Como puede apreciarse, la configuración emocional tiene forma de cuarto de luna. Esta forma viene determinada porque no hay ocurrencia de situaciones extremas en la valencia afectiva (tanto positiva como negativa) que presenten una baja activación, del mismo modo que cuando hay una extrema activación las situaciones no pueden ser neutras en valencia afectiva.

El principal atractivo de las propuestas dimensionales es que pueden dar explicación de un número ilimitado de estados emocionales y proporcionan un esquema para delimitar sus similitudes y diferencias entre las emociones.

El enfoque categorial

Este enfoque entiende que las emociones son *discretas*, partiendo de la existencia de características únicas y distintivas para cada categoría emocional. Entre estas características se han utilizado por ejemplo la correspondencia entre el tipo de *afrontamiento* -es decir, la movilización para la acción que producen-, y la propia forma emocional [26].

De entre todas las propuestas que se han utilizado para esta clasificación discretizada, es la orientación evolucionista [27], que propone la existencia de *emociones primarias* y *secundarias*, la más atractiva y más ampliamente aceptada. Las emociones primarias serían categorías emocionales primitivas -tanto filogenética como ontogenéticamente-[2] de carácter universal y a partir de las cuales se desarrollarán las demás emociones secundarias. Veamos algunas características de estos dos tipos de emociones:

- Según lo propuesto por Ekman, en los primeros momentos de la vida surgen las *emociones primarias*, entre las que se incluyen la sorpresa, el asco, el miedo, la alegría, la tristeza y la ira [28]. Cada una de estas emociones se corresponde con una función adaptativa primaria, de tal modo que, al menos en estratos superiores de la escala filogenética, éstas pueden observarse en individuos de diversas especies.

  Las emociones primarias poseen condiciones desencadenantes específicas y distintivas para cada una de ellas, un procesamiento cognitivo propio, una experiencia subjetiva característica, una comunicación no verbal distintiva (ver Figura 2.4) y un afrontamiento diferente.

[2]La *filogenia* hace referencia al origen, formación y desarrollo evolutivo de las especies, y la *ontogenia* a la formación y desarrollo individual de un organismo.

Figura 2.4: Las seis emociones primarias propuestas por Ekman

- Las *emociones secundarias*, también conocidas como sociales, morales o autoconscientes, corresponden con la culpa, la vergüenza, el orgullo, los celos, etc, y surjen en el individuo humano en torno a los 2 años y medio de edad o los 3 años [29], momento en el que su autoconcepto[3] se encuentra ya formado.

Tras esta introducción al concepto de emoción y la explicación de los diferentes enfoques para su estudio, procederemos a hablar de la síntesis de sonidos.

2.2 Síntesis de Sonidos

La síntesis de sonidos permite la obtención de sonidos a partir de medios no acústicos, lo cual aporta una gran libertad en cuanto a producción y perfilamiento de sonidos se refiere. Utilizando en sentido estricto la palabra síntesis, ésta consiste en la *generación* desde cero de un sonido, no en la *modificación* de uno preexistente [30].

A continuación haremos una breve explicación de los diferentes tipos y técnicas de síntesis que existen, para finalmente justificar nuestra elección de un tipo y una técnica concretos para realizar este proyecto.

2.2.1 Tipos de Síntesis y Técnicas de Síntesis

Atendiendo bien al proceso de síntesis, bien a la naturaleza del sintetizador, se pueden establecer dos clasificaciones del tipo de síntesis [30]:

1. *Analógica vs. Digital*: La síntesis analógica trabaja con conjuntos continuos de valores en forma de señales analógicas. Los sintetizadores analógicos utilizan dispositivos electrónicos capaces de producir este tipo de señales. En cambio, la síntesis digital trabaja en el dominio discreto. Debido a su versatilidad, este segundo tipo

[3] El *autoconcepto* es el estado mental que acontece cuando el niño se descubre como una parte separada del entorno que le rodea, con identidad propia, cuando puede entenderse a sí mismo como *yo*. Esto sucede alrededor de los dos años y es condición necesaria para que surjan emociones secundarias o sociales, que resultan de la comparación del *yo* con otras personas (la culpa, la vergüenza...). Hasta entonces, el niño sólo habrá experimentado emociones primarias, que regulan su conducta sin necesidad de una vinculación social. El lenguaje comienza a surgir por esta edad como herramienta social.

de síntesis permite el uso de infinitas técnicas de síntesis, además de permitir emular cualquier método de síntesis analógica.

2. *Hardware vs. Software*: Los sintetizadores digitales se dividen a su vez en sintetizadores por hardware y sintetizadores por software, basándose ambos en los mismos principios. Los sintetizadores por hardware tienen CPU, memoria, sistema operativo, etc, y contienen chips especializados en procesamiento de sonido (DSPs). Suelen incorporar un teclado musical y varios controles. Los sintetizadores por software, en cambio, son programas que aprovechan la tarjeta de sonido del ordenador.

En cuanto a las técnicas de síntesis [31], tenemos, entre las más utilizadas:

1. *Síntesis Aditiva.* Responde a cualquier método de síntesis por adición de diferentes sonidos en un sonido final. Dentro de esta técnica de síntesis se encuentran diferentes herramientas:
   (a) *Series de Fourier.* Se basa en el concepto de que cualquier onda puede ser completamente representada como una superposición de ondas sinusoidales complejas infinitesimalmente pequeñas, dada una amplitud y una fase, con valores variables de la frecuencia.
   (b) *Seguimiento de Pico y Síntesis de Modelado Espectral.* En el seguimiento de pico, la señal es representada como la suma de sinusoides que varían en amplitud y frecuencia a lo largo del tiempo.
2. *Síntesis Sustractiva.* Consiste en crear un sonido complejo e ir eliminando elementos de él, utilizando para ello filtros digitales que dan forma al espectro de frecuencias de la señal original. Entre los distintos tipos de Síntesis Sustractiva tenemos:
   (a) *Vocoders.* Una hilera de filtros paso-banda[4], distribuidos a lo largo del eje de frecuencias, es activada por una señal de banda ancha (por ejemplo, ruido blanco, onda cuadrada, tren de pulsos...). El valor de la ganancia en cada filtro es controlado y variado en el tiempo.
   (b) *Codificación Lineal Predictiva.* Esta técnica implementa un vocoder pero sustituye la hilera de filtros paso-banda por un único filtro más complejo que aproxima la respuesta en frecuencia de un instrumento (normalmente voz humana, o un instrumento de cavidades resonantes).
3. *Síntesis por Modulación.* Cualquier variación en el tiempo de una propiedad del sonido (amplitud, frecuencia...) se entiende como *modulación.*
4. *Síntesis por Modelos Físicos.* Los modelos físicos son modelos matemáticos que describen el comprtamiento de sistemas físicos. Se utilizan para modelizar el comportamiento del aire y elementos mecánicos en sistemas acústicos. Algunas de las técnicas más comunes son:
   (a) *Guías de Onda.* Basadas en la solución de la ecuación de ondas, describen el comportamiento de un tipo de onda en un medio concreto.

[4]Un filtro *paso-banda* es aquel que no altera las frecuencias que se encuentran dentro de un rango (banda) frecuencial determinado, eliminando (o atenuando) todas las demás. Por extensión, un filtro *paso-bajo* permite pasar las frecuencias bajas, y un filtro *paso-alto*, las altas.

(b) *Ecuaciones en diferencias.* En lugar de ecuaciones diferenciales se usan ecuaciones en diferencias, que dan una aproximación bastante ajustada a la realidad del comportamiento de estos sistemas.

5. *Síntesis Granular.* Esta técnica de síntesis está basada en una visión cuantificada del sonido, de tal modo que cada elemento (o *quasón*) que forma la totalidad de la expresión sonora está definido por envolventes de tono, frecuencia y amplitud. Así, la unión de estos quasones en distintas configuraciones pueden llegar a formar sonidos complejos y de gran riqueza acústica.

2.2.2 Parámetros sonoros

En el apartado 2.1.2 de este capítulo hemos hablado de los parámetros paralingüísticos de la comunicación oral. A continuación describiremos las cualidades primarias desde una perspectiva psicoacústica [32], con la intención de acercar al lector al análisis del sonido que hemos realizado en este trabajo.

1. *Tono*: es la sensación auditiva asociada a la frecuencia de la onda sonora. Cuando la onda responde a un comportamiento sinusoidal, la frecuencia y, por tanto, el tono de ésta es simple. Cuando la onda resulta de la suma de tonos puros o sinusoides puras de distintas frecuencias, el tono corresponde a la frecuencia del armónico fundamental. Se mide por tanto en Hertzios (Hz).

2. *Ritmo*: en música suele ser definido como el movimiento marcado por la sucesión regular de elementos débiles y fuertes. En el contexto de la expresión oral puede conceptualizarse como "el movimiento marcado por la sucesión de vocalizaciones fuertes y débiles que guían al receptor en intención pragmática". Se mide en bpm (beats per minute).

3. *Volumen*: hace referencia a la percepción subjetiva de la potencia de una onda sonora. Se mide en decibelios (dB).

4. *Timbre*: resulta de la combinación de varias ondas de sonido sinusoidales de distinta frecuencia que suenan de forma simultánea, y a distintas intensidades. La variación de la frecuencia y la intensidad de estas ondas determinan la variación del timbre. Para analizar un timbre se hace uso del espectro de frecuencias, que relaciona la intensidad de cada onda con su frecuencia determinada.

5. *Duración*: viene determinada por el tiempo en que una vocalización suena, y se mide en milisegundos (ms).

A continuación explicamos el concepto de envolvente, concepto de gran importancia en la evolución temporal de los sonidos que queremos generar.

2.2.3 Envolventes

El concepto de envolvente hace referencia al comportamiento de una característica del sonido a lo largo del tiempo, desde el comienzo del sonido hasta su desaparición [33]. El volumen, la frecuencia o el tono, por ejemplo, son características cuya evolución temporal puede ir definida por una envolvente. Toda envolvente tiene una serie de parámetros que

la definen. Pongamos un caso sencillo. El comportamiento temporal de la envolvente de volumen de la tecla de un piano, representado en la Figura 2.5, viene determinado por los siguientes parámetros:

- *Ataque*: determina el tiempo que el sonido parte del silencio hasta que alcanza su valor de volumen máximo.
- *Decaimiento*: define el tiempo en el que el sonido alcanza el nivel de volumen de la etapa de sostenido.
- *Sostenimiento*: especifica el volumen al que permanecerá el sonido mientras mantengamos presionada la tecla.
- *Relajación*: tiempo en el que el sonido se desvanece hasta alcanzar el silencio.

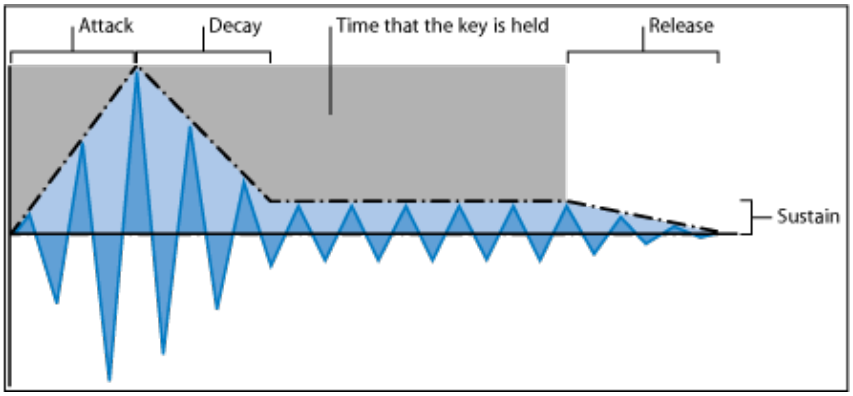


Figura 2.5: Parámetros de la Envolvente

Este ejemplo es extrapolable a otras características del sonido, como la frecuencia o el tono. En los siguientes capítulos veremos el uso que damos a este concepto en la síntesis de expresiones sonoras.

Capítulo 3

Plataformas y tecnologías utilizadas

En este capítulo pretendemos dar una visión comprensiva de las diferentes tecnologías que hemos utilizado en la realización del proyecto, y de las plataformas en las que las hemos implementado. La elección de cada una de éstas ha venido determinada por el diseño de nuestra aplicación y los objetivos que deseamos cumplir. Veamos cuáles son.

3.1 Maggie

En primer lugar, teniendo en cuenta que nuestro trabajo se encuentra dentro del marco de investigación de la interacción Humano-Robot, precisábamos de un robot de aspecto amigable pero con funcionalidades tecnológicas sofisticadas para implementar nuestra aplicación software. Maggie [34], el robot de RoboticsLab de la Universidad Carlos III de Madrid, cumplía este perfil. Os presentamos a Maggie:

"Sé que soy un poco bajita, apenas llego al metro y medio de altura, pero así puedo interaccionar mejor con los niños, que son igual de pequeños que yo", explica.

El diseño de Maggie va acorde con su funcionalidad y su objetivo principal: ser un robot social que interactúe directamente con los seres humanos. Sus formas redondeadas y su aspecto amable y amistoso (como vemos en la Figura 3.1) son las dos características que más resaltan a primera vista.

Figura 3.1: Maggie saludando

Maggie puede comunicarse mediante el habla, reconocer caras y seguir a su interlocutor mientras éste se mueve por por la habitación. Ofrece una apariencia antropomórfica y puede comunicarse a un nivel no verbal mediante el movimiento de sus párpados, de sus brazos y a través de la entonación. Además, en el pecho tiene un Tablet PC al cual van conectados un micrófono sin cables, que funciona por bluetooth, y dos altavoces. Este Tablet PC aporta información audiovisual y sirve como interfaz de interacción táctil para el usuario. En él Maggie inicia juegos y modos de interacción divertidos cuando interactúa con niños.

3.2 Juego "Kill the Croaches"

Además del robot, necesitábamos un entorno estimular con el cual Maggie pudiera interactuar y generar respuestas expresivas. Aprovechando el Tablet PC de Maggie, escogimos un juego en formato Flash desarrollado por uno de los investigadores del departamento de Ingeniería de Sistemas. En este juego se ve una circuitería -lo que serían los circuitos internos de Maggie-, que comienza a ser invadida por cucarachas que acceden por el ventilador de la placa, situado en la zona superior (como puede verse en la Figura 3.2). El jugador tiene que matarlas presionando con el dedo sobre ellas.

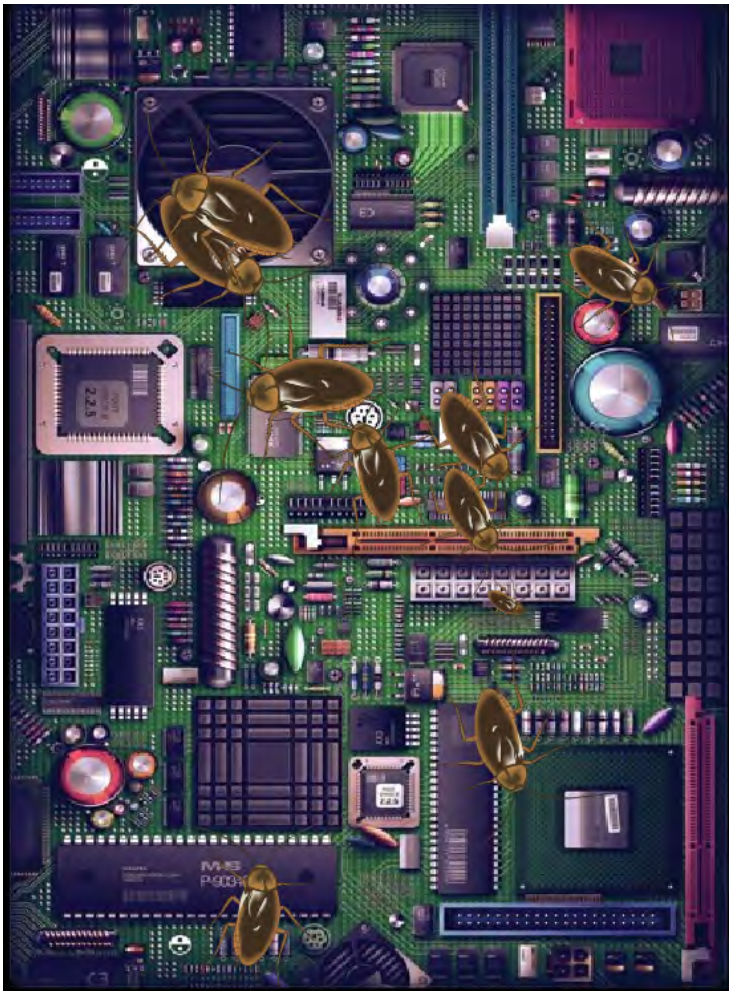


Figura 3.2: Vista del juego 'Kill the Croaches'

Una vez implementada la aplicación que hemos desarrollado, el juego cosistiría en lo siguiente: el objetivo es conseguir destruir un número de cucarachas igual o superior a $Nwin$ (prefijado previamente) para poder liberar a Maggie de la plaga. Si se consigue este ojetivo, Maggie se mostrará muy contenta. En cambio, si el número de cucarachas excede el valor $Nlose$ (otro valor prefijado), el juego terminará y Maggie se pondrá muy triste. Durante el juego, Maggie ofrecerá feedback auditivo al jugador sobre su rendimiento: expresará alegría si el jugador mata cucarachas, calma cuando no haya ninguna, y asco o miedo cuando el número de cucarachas vaya aumentando. Finalmente, expresará tristeza

cuando el número de cucarachas esté próximo a $Nlose$, o alcance éste valor. En el capítulo 4 se explicará esto con más detalle.

El juego tiene como variables de salida las siguientes:

- *Event*: encierra el valor asignado a un evento producido durante el juego. Toma el valor 0 cuando el evento es la aparición una nueva cucaracha, 1 cuando el jugador destruye a una, y 2 cuando el jugador pulsa sobre la superficie de la placa sin tocar ninguna cucaracha (con este valor se controlan los fallos del jugador).

- *Frequency*: determina la frecuencia de aparición de las cucarachas en Hz.

- *Number created*: su valor muestra el número de cucarachas que el juego ha creado durante la partida.

- *Number killed*: su valor muestra el número de cucarachas destruidas durante la partida. Con el número de cucarachas creadas y destruidas obtenemos el número de cucarachas n que en hay vivas en un momento determinado de la partida.

- *Time stamp*: determina el tiempo de partida en ms.

- *Life*: representa la vida del juego, y puede tomar valores enteros comprendidos entre 0 y 100.

Varias de estas variables de salida han sido utilizadas como estímulos o datos de entrada que recibe la aplicación software que hemos desarrollado. Como se puede suponer, la emisión de los valores de estas variables será continua y constante a lo largo de la partida, por lo que la aplicación recibirá estos datos en streaming. A continuación explicamos la plataforma que nos ha permitido recibir y manipular estos datos en streaming de una manera cómoda y flexible.

3.3 ROS

Según la página oficial de ROS (Robot Operating System) [35], este sistema software es una colección de herramientas y librerías que permiten simplificar la tarea de crear complejos y robustos comportamientos destinados a ser ejecutados por robots, a través de una amplia gama de plataformas.

Partiendo de la realidad de que ciertos problemas triviales a los ojos de los humanos son a menudo muy difíciles de implementar en software, ROS llega para facilitar el trabajo a todos aquellos que se dediquen al desarrollo de robots. ROS fue creado desde el principio para estimular el desarrollo de software en colaboración, y está diseñado especialmente para unir diferentes disciplinas dentro de la robótica en grupos de investigación conjuntos.

Figura 3.3: Logo de ROS

A continuación explicaremos brevemente el funcionamiento general de ROS, basándonos en los contenidos de su página oficial. Por simplicidad, obviaremos toda aquella información que no ha sido relevante para la realización de nuestro proyecto.

- *Nodo*: es un proceso que realiza la ejecución del código de cada tarea. Los nodos se combinan y comunican entre sí a través de *topics*, formando una red donde envían (publican en un topic) y reciben (se suscriben a un topic) información entre ellos. Esta configuración que huye de lo monolítico permite que, en el caso de que exista algún fallo, afecte únicamente a uno o pocos nodos, sin que el resto de éstos resulten afectados.

- *Tema o topic*: como decíamos, los topics son los canales a través de los cuales cada nodo se comunica con el resto, recibiendo datos de unos y enviando datos a otros. En general, los topics tienen semántica anónima en cuanto a publicación/suscripción, esto es, no saben de qué nodo proviene y a qué nodo va la información que contienen. De este modo, uno o múltiples nodos publican en un topic, y uno o múltiples nodos se suscriben a ese topic. Los topic están concebidos para transmitir información de forma unidireccional y en streaming.

- *Mensajes*: es la información publicada por los nodos en los topics. Un mensaje es una estructura simple de datos, que comprende un campo para cada tipo primitivo (int, float, string, boolean, etc) o array. La flexibilidad de los mensajes permite la anidación, de modo que un tipo concreto dentro de un mensaje puede ser otro mensaje creado previamente.

- *Paquete catkin*: éste es un paquete que contiene el código relativo a los mensajes y las tareas que se ejecutarán en los nodos, pero además contendrá un archivo llamado 'package.xml', que proporciona metainformación sobre el paquete, y otro archivo llamado 'CMakeList.txt', en el cual se especifica información relativa a la comunicación (por ejemplo, en este archivo habrá de especificarse los nombres de los mensajes que van a ser publicados).

- *Master*: el master de ROS proporciona servicios de nombramiento y registro en el sistema. Realiza un seguimiento de los nodos que publican y suscriben en topics. El papel del master es permitir que cada nodo localice a aquellos nodos con los que interactuará. Para correr el master se utiliza el comando *roscore*, que carga el master junto con otros componentes esenciales.

- *Roscore*: es una colección de nodos y programas que son fundamentales para el funcionamiento de un sistema basado en ROS. Para que exista comunicación entre los nodos creados por el investigador, es necesario que roscore esté corriendo en una terminal aparte.

- *Rosrun*: éste es un comando que permite correr un nodo dentro de un paquete especificado, sin necesidad de saber el PATH[1] del paquete.

- *Roslaunch*: es un comando que corre los nodos que se han especificado en un archivo de extensión .launch (escrito en xml), pudiendo concretarse agrupaciones y mapeos entre ellos.

ROS soporta Python y *C++*, de modo que los códigos que posteriormente serán ejecutados pueden estar escritos en uno de estos dos lenguajes.

[1]En informática, el *PATH* es una variable de entorno en la que se especifica las rutas en las cuales el intérprete de comandos debe buscar los programas a ejecutar.

3.4 Python

Python [36] es un lenguaje de programación interpretado cuya sintaxis se aleja de otros lenguajes de uso frecuente como Java o C por su legibilidad y simplicidad. Es un lenguaje multiparadigma, pues soporta orientación a objetos, programación imperativa y, en menor medida, programación funcional. El tipado dinámico de Python permite asignar a una variable valores de distinto tipo.

Figura 3.4: Logo de Python

Como decíamos en el anterior apartado, es un lenguaje soportado por ROS. Además, Python cuenta con una muy elevada cantidad de librerías desarrolladas desde el marco del software libre que permiten hacer casi cualquier cosa. En referencia a esto, ha sido necesaria la descarga de 'Mido', una librería que trabaja con mensajes MIDI y puertos. Más adelante explicaremos a grandes rasgos el protocolo midi, y en concreto, el uso que hemos hecho de esta librería.

3.5 Ableton Live Suite 9

3.5.1 Elección para la Síntesis de Sonidos

Antes de proceder a explicar el tipo de herramientas que nos ofrece Ableton Live [37], justificaremos por qué elegimos este DAW [2] para realizar la síntesis de sonidos. En primer lugar, llegamos a la conclusión de que un sintetizador por *software* sería el más apropiado para nuestro trabajo, antes que uno que funcionase por *hardware*, principalmente debido a las siguientes razones [30]:

- *Economía* para la experimentación: muchos programas *freeware* o *shareware* ofrecen posibilidades de síntesis innovadoras, experimentales, que no han sido implementadas por ningún fabricante en un sintetizador por *hardware*. Esto nos abre un abanico más amplio de oportunidades.
- *Flexibilidad* en el método: existen muy variadas técnicas y algoritmos para la síntesis digital, y la mayoría de los dispositivos *hardware* sólo incorpora una de ellas. La síntesis por *software* permite además implementar cualquier método o algoritmo, que funcionará en cualquier ordenador independientemente de la tarjeta de sonido que tenga, siempre que el ordenador disponga de la potenia suficiente.
- *Combinación de métodos*: posibilidad de encadenar muchos de estos programas para generar arquitecturas complejas, lo cual es inalcanzable para un sintetizador *hardware*.
- *Existencia de entornos específicos*: existen entornos y lenguajes de programación especialmente enfocados a la síntesis que nos permiten crear nuestros propios sintetizadores sin ningún límite (MAX, Supercollider, Pure Data...).

[2] Un DAW o Estación de Trabajo de Audio Digital (*Digital Audio Workstation*, en inglés) es un *software* que permite la grabación, edición y producción de archivos de audio, como canciones o grabaciones de voz. Incluso ciertos DAW, como es el caso de Ableton Live, incluyen plugins y aplicaciones de síntesis de sonido.

- *Posibilidad de procesado*: la flexibilidad ha roto la barrera entre síntesis y procesado, que puede realizarse en tiempo real.

Ciertas desventajas de la síntesis por *software*, como la latencia o el ruido, no han sido significativas en las pruebas preliminares realizadas, de modo que optamos trabajar con este tipo de síntesis.

El siguiente paso a dar fue decidir qué programa utilizar. Como explicábamos antes, existen entornos creados específicamente para la síntesis de sonido, sin embargo, optamos por usar un DAW.

La principal ventaja de utilizar un DAW en lugar de un entorno específico es su fácil empleo, al alcance de cualquier usuario, ya que las interfaces de éstos suelen ser atractivas para el usuario y la disposición de los elementos para generar y manipular ondas es sencilla y ordenada. El manejo es intuitivo y los resultados pueden llegar a ser de calidad profesional.

Por otro lado, en el caso de los entornos específicos de síntesis, aunque la potencia de generación es muy elevada, no compensa la complejidad de su uso y el hecho de contar con un lenguaje de programación propio. Sólo tras un aprendizaje relativamente extenso de la sintaxis de su lenguaje pueden llegar a la realizarse trabajos con la calidad deseada.

Una vez decidimos utilizar un DAW, exploramos las opciones disponibles y finalmente nos decantamos por Ableton Live. Dejando a un lado las numerosas herramientas que este programa permite más allá de la síntesis de sonido, nos centraremos en aquellas que hemos utilizado. Ableton Live 9 cuenta con un instrumento *software* llamado Operator, que será explicado en el siguiente apartado con todo detalle.

3.5.2 Operator

Operator[3] es una herramienta de síntesis que combina la síntesis sustractiva al estilo analógico y síntesis de modulación de frecuencia. Muy sencillo de usar, Operator permite un abanico de posibilidades en la creación de sonidos, contando además con una colección de presets (sonidos pre-programados) manipulables a partir de los cuales se pueden generar sonidos muy interesantes. La apariencia de Operator es la representada en la Figura 3.5.

Figura 3.5: Vista general de Operator

Vemos que pueden combinarse hasta cuatro osciladores en distintas arquitecturas de modulación. En la parte izquierda del plugin podemos observar los controles de las ondas A, B, C y D: los desfases tonales *Coarse* (desfase grueso) y *Fine* (desfase fino) y el limitador

[3]Esta herramienta mantiene los nombres de todos sus parámetros en inglés, por lo que cuando nos refiramos a ellos indicaremos a pie de página su traducción al español.

de volumen *Level*. En la parte central observamos un panel que nos permite manipular la envolvente de volumen de cada oscilador (en este caso, del A), accediendo a una gran cantidad de parámetros de onda.

Si queremos personalizar el tipo de modulación que tendrán los cuatro osciladores y la configuración, podemos elegir entre distintas opciones de algoritmos de modulación, con la posibilidad de modificar otros parámetros. Esto podemos verlo en el panel central de la Figura 3.6:

Figura 3.6: Algoritmos de Modulación

Podemos generar el tipo de onda que queramos editando los armónicos de la onda con su intensidad correspondiente (Figura 3.7).

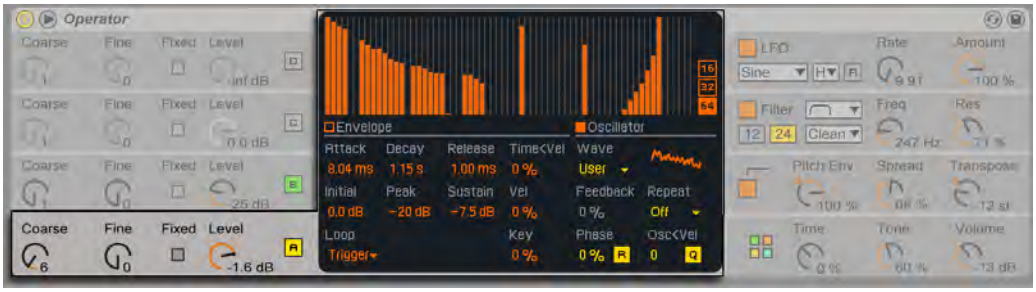


Figura 3.7: Oscilador

Operator también cuenta con una serie de filtros, un LFO (Oscilador de Baja Frecuencia, del inglés *Low-Frequency Oscillator*), una envolvente de timbre y otra de tono. El LFO permite crear un pulso rítmico que describe el comportamiento de un tipo de onda concreto (senoidal, triangular, diente de sierra, cuadrada, aleatoria...) y asignarlo a una característica del sonido, como el volumen, el panorámico o el tono; o asignarlo a un filtro paso-bajo, a una envolvente de timbre, etc.

Vemos en la Figura 3.8 que el LFO de Operator cuenta con controles como *Ratio (Rate)*, que determina la frecuencia de variación de la característica de sonido asignada, y *Amount*, que controla la amplitud de la variación de tal característica. Además, el LFO de Operator cuenta con su propia envolvente, lo que permite una personalización mayor de la onda. Otra potente herramienta es el filtro de frecuencias (Figura 3.9), que permite seleccionar entre una gran variedad de tipos (paso-alto, paso-bajo, paso-banda...) y que cuenta, también, con su propia envolvente. Finalmente, Operator tiene una envolvente de tono (Figura 3.10) que además de parametrizarse, se controla por medio de tres selectores:

- *Envolvente de Tono (Pitch Envelope)*: con él seleccionamos el porcentaje de control que tiene la envolvente de tono sobre la onda.

- *Apertura (Spread)*: por medio de la duplicación y detuning crea un coro estéreo.
- *Transposición de tono (Transpose)*: controla el tono global del sonido que estemos sintetizando.

Figura 3.8: LFO

Figura 3.9: Filtro

Figura 3.10: Envolvente de Tono

Este último instrumento nos ha permitido elaborar distintas entonaciones adecuadas a cada expresión sonora. En los apartados que vienen a continuación explicaremos otra serie de elementos que, junto con Operator, nos han permitido manejar cada característica del sonido con una libertad casi total.

3.5.3 MIDI Effects

MIDI es un estándar tecnológico que describe un protocolo, que fue creado para la interacción entre sintetizadores digitales y otros dispositivos de música digitales. Los efectos MIDI se implementan sobre sonidos que se rigen por este protocolo. Más adelante explicaremos la tecnología MIDI con más detenimiento.

Arpeggiator

Este efecto MIDI es un clásico de la música de los 80, que convierte notas MIDI individuales en acordes con un patrón rítmico determinado. La interfaz de este efecto se muestra en Ableton según la Figura 3.11.

A continuación explicamos los controles que hemos utilizado para nuestro proyecto:

- *Ratio (Rate)*: determina el tiempo entre repetición y repetición de la nota, de tal modo que a mayor ratio, más separación entre notas. Puede tomar valores entre 10 ms y 1 s.
- *Repeticiones (Repeats)*: configura el número de repeticiones de la nota, y su rango va de 1 a 16, en pasos de entero, y tiene la opción de infinitas repeticiones (inf).
- *Distancia (Distance)*: determina el rango de variación de notas en el arpeggio en pasos (steps, st, cada uno de éstos corresponde a un semitono). Puede tomar valores de -24 st a 24 st, es decir, hasta un total de cuatro octavas.
- *Puerta (Gate)*: configura la longitud de las notas como un porcentaje del valor actual de *Ratio*. Tiene un rango entre 0 % y 200 %. Valores por encima del 100 % generarán legatos.

Figura 3.11: Arpeggiator

Note Length

La función de este efecto (cuya interfaz se muestra en la Figura 3.12) es alterar las longitudes de las notas y los silencios MIDI de entrada. El control utilizado en nuestro trabajo ha sido *Longitud de nota (Length)*, cuyos valores van de 10 ms a 60 s.

Figura 3.12: Note Length

Random

Este efecto (Figura 3.13) genera una respuesta tonal aleatoria dentro de un rango de valores prefijado. El valor aleatorio que determina el tono es creado por dos variables: el control *Elecciones (Choices)* define el número de diferentes notas aleatorias, en un rango de 1 a 24 (hasta dos octavas). El control *Scale* es multiplicado por el valor fijado del control *Choices*, y el resultado determina el rango posible de tonos.

Figura 3.13: Random

3.5.4 Instrument Rack

Un Rack es una herramienta flexible para trabajar con efectos, plugins e instrumentos en una cadena de dispositivos de pista. Puede ser utilizado para construir procesadores de señal complejos, instrumentos de comportamiento dinámico, sintetizadores, etc. La vista general de esta herramienta (Figura 3.11), con Operator y otros efectos ensamblados, permite observar el alcance de su uso.

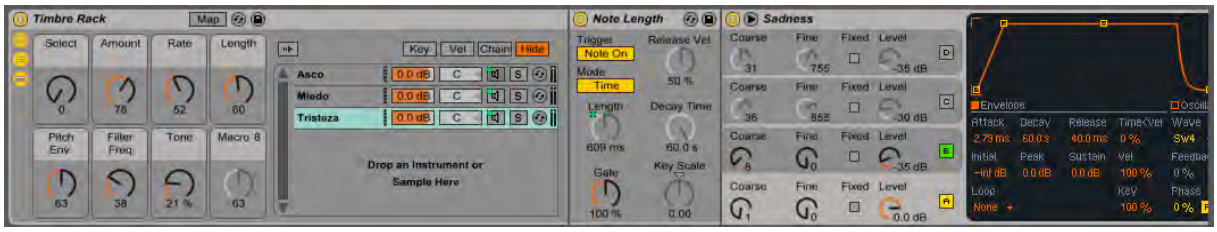


Figura 3.11: Instrument Rack

La herramienta *Instrument Rack* se divide en dos partes:

- *Controles Generales*: esta área está situada a la izquierda de la imagen. Está compuesta por ocho controles generales sin asignación inicial. Por medio del *Modo de Mapeo Macro (Macro Map Mode)* (opción situada en la cabecera, bajo el nombre de *Map*), puede realizarse una asignación a cada control general de cualquier control o selector de efectos MIDI y Operator, siendo ilimitado el número de selectores que pueden ser asignados a cada control general. De este modo puede realizarse una parametrización dinámica (concepto que será explicado en el próximo capítulo) desde el control general de cualquier selector de la cadena.

- *Lista de Cadenas*: en esta área, contigua a la anterior según la Figura 3.11, pueden crearse cadenas de efectos, por lo que en el mismo Instrument Rack puede haber varios sonidos distintos e independientes entre sí, cada uno asociado a una cadena.

3.5.5 Vista de Tracks

El sistema que hemos usado en Ableton para ordenar y utilizar los sonidos que hemos generado es modular. Ableton nos ha permitido hacer esto fácilmente, a través de tracks, en cada cual puede colocarse un Instrument Rack con sus cadenas de efectos.

Como puede verse en la Figura 3.12, hemos creado dos tracks: *Placer* y *Displacer*. Cada track contiene las emociones afines a estas categorías. Cada track cuenta con un botón de encendido y apagado, lo que nos ha servido para silenciar aquellos que durante la expresión vocal no están en uso.

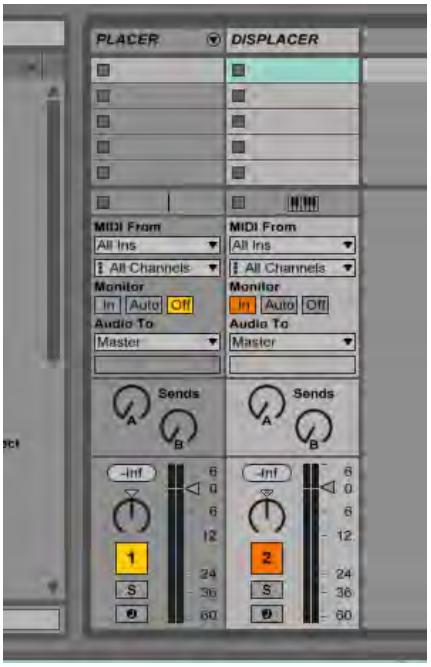


Figura 3.12: Vista de Tracks

3.5.6 MIDI Map Mode

El modo de mapeo MIDI permite crear rutas MIDI desde el controlador (un teclado MIDI, otro software..., en nuestro caso, el nodo en ROS) hasta el propio Ableton. Como se puede ver en la Figura 3.13, aquellas áreas coloreadas de azul pueden ser mapeadas. El mapeo consiste en la asignación de un canal y un valor de control. Vemos que los Controles Generales del Instrument Rack han sido mapeados, así como los botones de encendido y apagado de cada track.

Como hemos dicho, los tracks de Ableton Live reciben los mensajes MIDI procedentes de la aplicación en ROS. Expliquemos brevemente en qué consiste este tipo de comunicación.

3.6 Protocolo MIDI

Las siglas MIDI corresponden a *Musical Instrument Digital Interface*, o Interfaz Digital de Instrumento Musical, en español. MIDI es un estándar tecnológico que describe un protocolo, el cual fue creado para la interacción entre sintetizadores digitales y otros dispositivos de música digitales, aunque actualmente permite la comunicación entre instrumentos, ordenadores, TV, tablets, smartphones, etc. Esto es debido a que es un protocolo de latencia mínima.

El protocolo MIDI es un bloque de "comandos musicales", cada uno de los cuales consiste en un byte de estado y en uno o dos bytes de datos. Si unimos dos de estos bytes de distinto tipo obtenemos un mensaje. El dispositivo que emite mensajes se llama controlador, y aquel que recibe estos mensajes es el módulo de sonido.

Existen diferentes tipos de mensaje MIDI, cada uno relacionado con una acción musical específica. Para entender la función de los distintos tipos de mensajes hemos utilizado,

Figura 3.13: Modo de mapeo MIDI

haremos uso de las especificaciones de la librería de Python con la que hemos trabajado para generar mensajes MIDI, llamada Mido [38]. Éstos son:

1. *Acción*: NOTE ON (envía una señal que activa un sonido asignado), NOTE OFF (envía una señal que detiene la ejecución el sonido), CONTROL (envía una señal de control sobre un elemento que puede ser mapeado, como un interruptor o un selector).

2. *Canal*: medio a través del cual se realiza la comunicación mediante mensajes. Tiene un rango de valores de 0...15, de modo que un único controlador puede llegar a conectarse con 16 módulos de sonido.

3. *Nota*: valor asignado a cada nota musical, dentro de un rango entre 0...127.

4. *Velocidad*: valor asignado a la intensidad de ejecución de la nota, dentro de un rango entre 0...127.

5. *Control*: es un tipo de parámetro cuyo valor se asigna a un elemento mapeable, pudiendo tener hasta 128 elementos bajo control dentro de un mismo canal.

6. *Valor*: valor asignado a la posición del selector o al estado del interruptor, dentro de un rango entre 0...127.

Capítulo 4

Sistema propuesto

Como recordamos, los objetivos de este trabajo son los siguientes:

- Desarrollar una aplicación software orientada a la interacción Humano-Robot, que capte estímulos del entorno, los procese y transforme en patrones de expresión sonora no verbal en tiempo real.
- Generar mediante síntesis de sonido un grupo de NLUs que expresen emociones, a través de la tecnología Operator del *software* Ableton Live 9 Suite.
- Implementar el sistema de interacción (juego + aplicación software + sonidos generados por síntesis) en Maggie.

En este capítulo haremos una explicación del trabajo que hemos llevado a cabo para cumplir estos objetivos. Para ello, hemos dividido el contenido en dos bloques generales.

En el primer bloque haremos una descripción del sistema propuesto, aportando una visión general de la aplicación que permita entender de manera visual y directa el proceso que se realiza, desde que el jugador pulsa la pantalla hasta que se oye la expresión sonora. A continuación explicaremos, a partir de un diagrama de estados, las condiciones del juego que causan transiciones entre estados emocionales. Después, justificaremos el enfoque que hemos escogido para el estudio de las emociones y qué emociones forman el rango de respuesta de nuestra aplicación. Finalmente, haremos un breve análisis de los elementos paralingüísticos que hemos tenido en cuenta para sintetizar nuestras expresiones sonoras.

En el segundo bloque explicaremos la implementación de nuestro sistema, analizando la arquitectura de la aplicación y su comportamiento. Por último, daremos una explicación exhaustiva del proceso y los resultados de la síntesis.

4.1 Descripción del sistema propuesto

4.1.1 Visión general

Explicado de forma sencilla, la función de nuestra aplicación es transformar los datos que recibe del juego en expresiones sonoras que sean acordes a lo que está sucediendo en el juego. Para esto, hemos dividido las funciones de la aplicación en tres subsistemas, que a su vez se componen de varios módulos que ejecutan acciones específicas. Veamos cómo funciona exactamente:

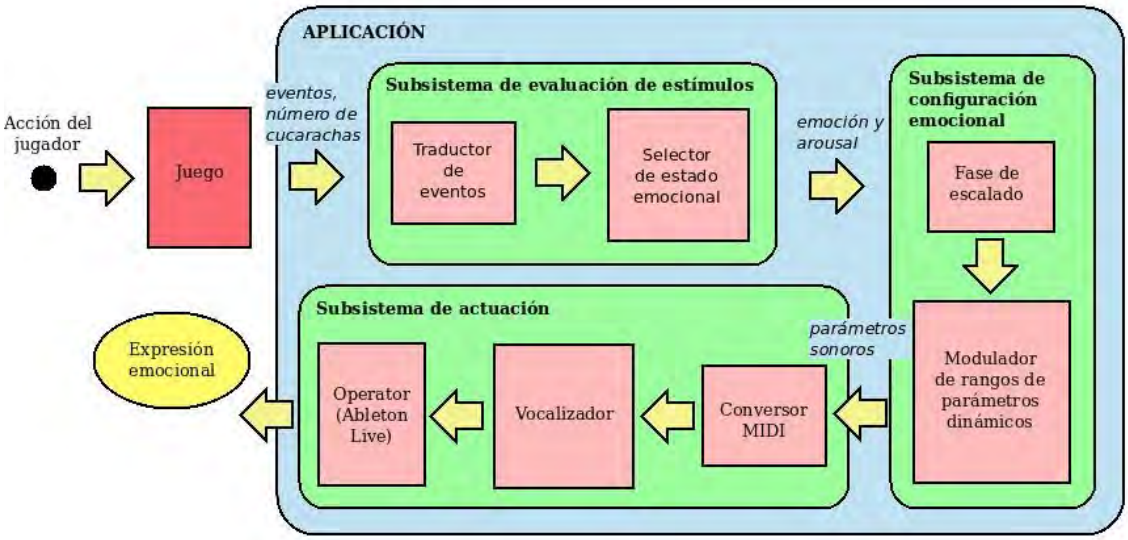


Figura 4.1: Diseño general de la aplicación

- El *Subsistema de evaluación de estímulos* imita al sistema de análisis propuesto desde la Psicología de la Emoción. Actúa como primer filtro en la detección y desencadenamiento emocional, discriminando lo que tiene relevancia emocional de lo que no la tiene y dándole un grado de intensidad. Este sistema está formado por:
  - Un *Traductor de eventos*, que transforma los datos de entrada al sistema (las variables de salida del juego 'Kill the Croaches') en valores de activación emocional (arousal) y da un significado al valor de la variable *event*.
  - Un *Selector de estado emocional*, que interpreta los valores proporcionados por el Traductor de eventos y los asigna a una emoción.
- El *Sistema de configuración emocional* realiza las operaciones necesarias para, a partir de la emoción y el nivel de arousal, configurar los parámetros sonoros de la vocalización que se va a realizar. Este sistema está formado por:
  - Una *Fase de escalado*, que traduce los valores de arousal (entre 0 y 1) en valores MIDI (entre 0 y 127).
  - Un *Modulador de rangos de parámetros dinámicos* que, alimentado con los valores de la fase previa, establece los rangos de valores que podrá tomar cada uno de los parámetros sonoros de las envolventes y otros elementos de la vocalización.
- El *Sistema de actuación* pone en funcionamiento la generación de la expresión sonora de la emoción. Este sistema está formado por:
  - Un *Conversor MIDI*, que dispone los valores de salida del Modulador de rangos en forma de mensajes MIDI.
  - Un *Vocalizador*, que efectúa el algoritmo específico que crea la "articulación" de la respuesta vocal.
  - *Operator (Ableton Live)*, que toma la articulación del Vocalizador y genera el sonido acorde a ella.

4.1.2 Diagrama de estados

En el diseño de nuestra aplicación, el *Selector de estado emocional* se representa en la Figura 4.2 como un diagrama de estados. CALMA funciona como el estado en el que se inicia la aplicación, perviamente a la aparición de las cucarachas. N_{win} hace referencia al número de cucarachas que hay que destruir para que la partida se dé por ganada, y N_{lose} al número de cucarachas vivas con el que el jugador pierda la partida. Estos valores pueden ser prefijados antes de comenzar a jugar la partida. n es el número de cucarachas vivas en un instante de la partida, n_0 el número de cucarachas vivas en el instante previo a un evento. Las variables *number_killed*, *number_created* y *event* son datos de salida del juego, como ya vimos.

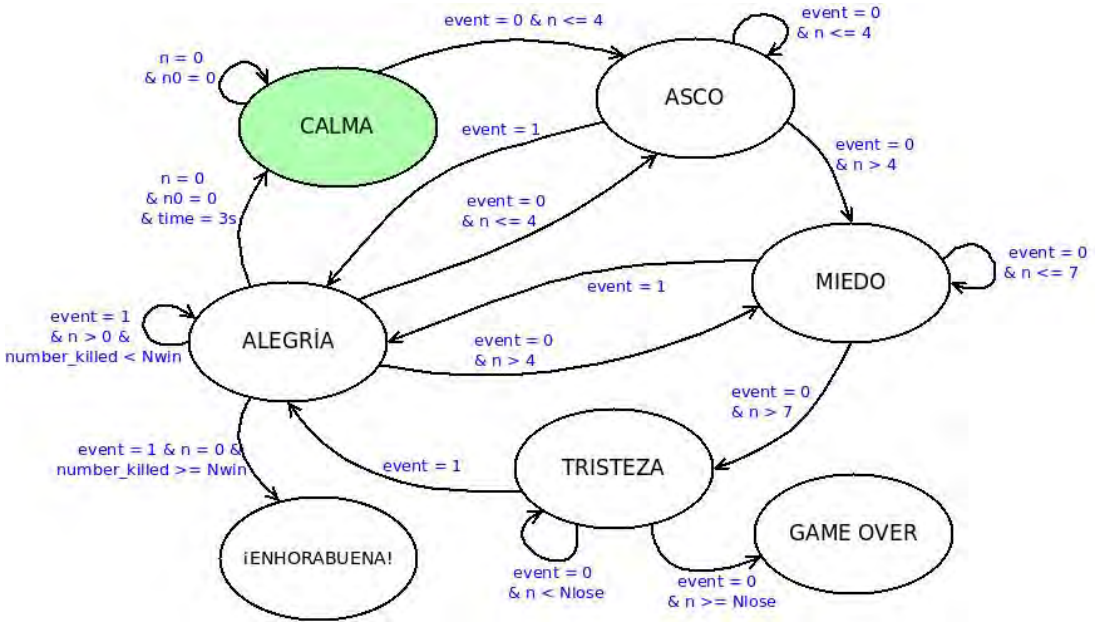


Figura 4.2: Diagrama de Estados

En la Tabla 4.1 se muestran los valores de las salidas *state* (la emoción) y *arousal* (intensidad de la emoción) de nuestro sistema en cada estado. La salida *arousal* puede tomar valores entre 0 y 1. Como esta salida depende n, la hemos normalizado en función del rango de valores que puede tomar n en cada estado.

4.1.3 Enfoque de estudio y elección de las emociones

En este apartado hablaremos de la configuración emocional de nuestro sistema, que se basa en dos consideraciones fundamentales:

1. La adopción del enfoque categorial en la comprensión de las emociones, si bien haciendo uso de un continuo de arousal propio del enfoque dimensional.

2. La elección de expresiones emocionales básicas o primarias y determinación de sus características sónicas a partir del análisis paralingüístico.

State	Arousal
CALMA	0
ASCO	$\frac{n}{4}$
MIEDO	$\frac{n-5}{3}$
ALEGRÍA	$\frac{n-8}{max\{N_{lose},N_{win}\}-7}$
TRISTEZA	$\frac{n-1}{N_{lose}}$
¡ENHORABUENA!	1
GAME OVER	1

Tabla 4.1: Salidas de la máquina de estados

Adopción de características del enfoque categorial

Como decíamos en el capítulo 2, en las últimas décadas se optó por combinar el enfoque categorial y dimensional en el estudio de la emoción, dando lugar a una explicación más completa de ésta como concepto y de cada una de las emociones por separado.

De este modo, y de acuerdo con el enfoque categorial, conceptualmente entendemos cada una de las expresiones emocionales que hemos estudiado como independientes porque, primero, responden a características ambientales diferentes, y segundo, porque cada una tiene unas características sonoras específicas y distintivas, que la hacen diferente a las demás.

Por otra parte, y de acuerdo con el enfoque dimensional, cada expresión emocional posee diferentes grados de intensidad expresiva a lo largo de un continuo, en función de la activación generada por estímulos externos. A este respecto, no nos ha parecido necesario utilizar el concepto de valencia en nuestro modelo, debido a que cada emoción es una categoría cerrada o estado, a la cual se llega bajo condiciones estimulares concretas, como veremos más adelante.

Elección de expresiones emocionales básicas

En cuanto a la elección del tipo de emoción, hemos optado por el estudio de las emociones primarias, antes que secundarias, por varias razones:

- La expresión de emociones primarias, de acuerdo con la perspectiva evolucionista, es universal, ya que éstas surjen en etapas tempranas en el desarrollo filogenético y ontogenético, y es, por tanto, poco o nada dependiente de la cultura o la localización geográfica del individuo.
- Al aparecer en etapas previas al desarrollo del autoconcepto del individuo, estas emociones no tienen por qué estar vinculadas al lenguaje, de modo que la expresión vocal no verbal sería suficiente para expresar cada una de éstas.
- Como corolario del anterior punto, la expresión vocal no verbal de una emoción primaria, como el miedo, es mucho más sencilla de generar en términos de síntesis sonora que la expresión de una emoción secundaria, como por ejemplo, la vergüenza.

Tomando como referencia las conclusiones de los estudios sobre emociones desarrollados por Paul Ekman [39], entre las emociones que emergen en los primeros momentos de la vida se incluyen la *alegría*, el *asco*, el *miedo*, la *sorpresa*, la *tristeza* y la *ira*. Para nuestro estudio hemos escogido, de entre éstas, la alegría, el asco, el miedo y la tristeza, y un estado emocional base, la calma. Haremos una breve definición de cada una y situaremos su significado en el contexto del juego:

Alegría

Esta emoción básica genera un sentimiento positivo como resultado de: a) la atenuación de malestar; b) la consecución de alguna meta u objetivo deseado (cuyo logro no tiene que ser necesariamente esperado); c) una experiencia estética. En nuestro estudio hemos simplificado el contexto de aparición de esta emoción únicamente al apartado a), por lo que Maggie expresa alegría cuando el número de cucarachas (motivo de su malestar) disminuye debido a la acción del jugador. Cuando el número de cucarachas llega a un valor N_{win}, Maggie expresará su máximo de alegría y el juego termina con ¡ENHORABUENA!

Asco

En sentido más general, el término *asco* define la respuesta emocional causada por la repugnancia o impresión desagradable hacia algo, sea un ser vivo, una sustancia, una conducta o ciertos valores culturales.

La expresión sonora de asco aparece cuando en el interior del cuerpo de Maggie comienzan a surgir cucarachas que corretean. Con un número determinado de cucarachas, este asco se convertirá en miedo.

Miedo

El *miedo* puede conceptuarse como una emoción producida por un peligro presente e inminente, y que se encuentra muy ligada al estímulo que lo genera.

En nuestro trabajo, el estímulo generador de la respuesta de aversión es la potencial amenaza que supone para Maggie el número de cucarachas en su interior.

Tristeza

La tristeza es una emoción negativa caracterizada por un decaimiento en el estado de ánimo habitual, que se acompaña de una reducción significativa del nivel de actividad cognitiva y conductual, así como de la activación fisiológica. El sentimiento de tristeza, es decir, la experiencia subjetiva, emerge ante situaciones que suponen a) la pérdida de una meta valiosa; b) el planteamiento de una contingencia aversiva inescapable.

En el caso que nos ocupa, entendemos la expresión de tristeza como la consecuencia de una contingencia aversiva inescapable para Maggie, que sucede cuando el número de cucarachas ha sobrepasado el límite de tolerancia, y el juego termina con GAME OVER.

Calma

La *calma* no forma parte del grupo de emociones primarias descritas por Ekman, pero nos ha parecido necesario incluirla en nuestro trabajo como un estado base positivo, del

cual parte la interacción del robot con el entorno.

A grandes rasgos, y en términos fisiológicos, la calma podría conceptualizarse como una disposición del afecto del individuo cuando éste está en homeostasis[1], del cual parte y al cual tiende en condiciones ambientales y mentales normales. Un estado opuesto a la calma tendría como característica principal una activación fisiológica elevada.

Debido a que no existe actualmente mucha literatura de fácil acceso sobre la expresión vocal de la calma, hemos decidido libremente dar nuestra interpretación a esta expresión. Los motivos por los cuales esta información documental es escasa podrían ser:

- Según la perspectiva evolucionista de las emociones, cada una de las emociones primarias se corresponde con una función adaptativa, y su comunicación viene determinada por su utilidad en términos de supervivencia directa (en el caso del miedo o el asco), de repercusión positiva con el grupo social (en el caso de la alegría), de obtención de ayuda emocional (en el caso de la tristeza) o de establecimiento de límites en una relación (en el caso de la ira). En lo que acontece a la expresión de calma como herramienta comunicativa, no respondería a ningún tipo de necesidad como los descritos anteriormente, al menos en estadios evolutivos más primarios de la especie humana, por lo que su expresión vocal no habría sido evolutivamente definida.

- La expresión de la calma, lejos de ser universal (como las emociones primarias), respondería en gran parte a claves culturales e históricas, siendo por tanto amplísimo el abanico de vocalizaciones que podrían denotar un estado de calma.

Teniendo en cuenta estas consideraciones, hemos optado por elaborar una expresión vocal de este estado muy concreta y específica, consistente en la ejecución de breves vocalizaciones, de frecuencia, longitud y tono variables, similar al "parloteo" que hace un periquito cuando está en una situación distendida. Esta expresión vocal será debidamente explicada en el próximo apartado.

4.1.4 Análisis de elementos paralingüísticos

Considerando los trabajos de Poyatos sobre los parámetros sonoros de las expresiones orales, mostramos en la Tabla 4.2 una aproximación que nos ayudó a enfocar la síntesis sonora de cada expresión emocional y nos guió durante tal proceso. Debe entenderse que esta tabla ofrece una descripción orientativa, y de ningún modo exhaustiva, de las cualidades primarias de cada expresión.

[1] La *homeostasis* es la función autorreguladora del organismo por medio de la cual se mantiene una constancia relativa en la composición y propiedades de su medio interno.

Emoción	Tono	Inflexión[a]	Ritmo	Volumen	Timbre[b]
Alegría	Alto	↗↘	Elevado	Alto	Medio y Medio-Redondeado
Asco	Medio	↗	Arrítmico	Medio	Cerrado y Redondeado
Miedo	Muy Alto	↗↗	Arrítmico	Muy Alto	Abierto y Áspero
Tristeza	de Alto a Bajo	↘→↘	Cadencia	Alto	Cerrado y Áspero
Calma	Medio	Aleatoria	Aleatorio	Medio-Bajo	Cerrado y Redondeado

Tabla 4.2: Cualidades primarias de las expresiones emocionales

[a]Nota: ↗≡ inflexión ascendente; ↘≡ inflexión descendente; →≡ inflexión mantenida; ↗↗≡ inflexión ascendente elevada.

[b]Un timbre *áspero* es generado por una onda de tipo diente de sierra, cuadrada o triangular. Un timbre *redondeado* puede ser generado por una onda sinusoidal.

4.2 Implementación del sistema

Para mejorar la comprensión del sistema propuesto, este apartado ha sido dividido en dos bloques generales:

1. En el primero explicaremos la arquitectura de nuestra aplicación y la funcionalidad de sus clases más importantes, entre las que se encuentra la configuración y control de los parámetros dinámicos que completan la caracterización de nuestras expresiones emocionales.

2. En el segundo bloque ofreceremos una visión pormenorizada de la síntesis de expresiones emocionales que hemos realizado.

4.2.1 Arquitectura y comportamiento de la aplicación

El diagrama de clases[2] que presentamos a continuación corresponde al diseño de la aplicación que hemos visto al comienzo del capítulo. En él pueden verse las clases que hemos creado (aquellas en las que se detalla sus datos y métodos)[3] y otras que pertenecen a ROS o a otras librerías. Como se observa en la Figura 4.3, se ha pretendido una configuración jerárquica de visualización, que permite mostrar las capas de funcionamiento:

- En la capa superior, las clases *Publisher* y *Subscriber* de ROS generan la comunicación entre nodos.

- La segunda capa está formada por los nodos *SocketInterface* -que conecta el juego 'Kill the Croaches' con la aplicación-, *ArousalGenerator*, *EmotionGenerator* y *ActionGenerator*, que realizan la tarea comunicativa. Estos nodos crean objetos de la clase *Rate*, otra clase de ROS que regula el ratio de entrada y salida de datos en cada nodo.

[2]Un diagrama de clases es, en UML (Unified Modeling Language en inglés), un tipo de diagrama estático que describe la estructura de un sistema mostrando sus clases, en programación orientada a objetos.

[3]Por motivos de espacio, no se han detallado los datos y los métodos de aquellas clases que no han sido desarrolladas en este trabajo.

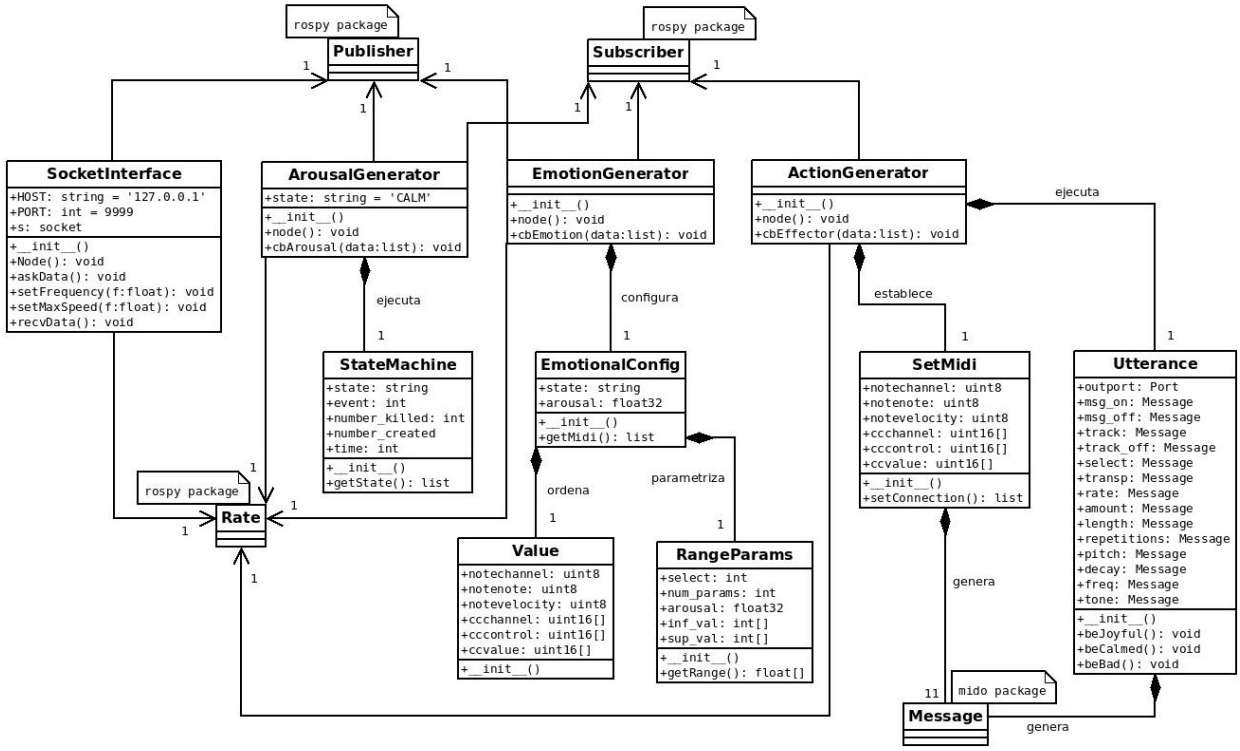


Figura 4.3: Diagrama de Clases de la aplicación

- Las clases de la tercera y cuarta capa contienen funciones más específicas relacionadas con la lógica del negocio de nuestra aplicación, cada una enfocada en una tarea:
  - La clase *StateMachine* contiene el algoritmo que define las condiciones de cada expresión emocional y las transiciones entre cada una de ellas. Este algoritmo es la máquina de estados que hemos explicado al inicio del capítulo.
  - La clase *RangeParams* recibe el arousal del estado actual, y los valores máximos y mínimos experimentales que comprenden el rango de cada parámetro dinámico configurado en Ableton. Mediante un procedimiento de escalado lineal, transforma el valor del arousal (entre 0 y 1) en valores comprendidos entre 0 y 127, que e el siguiente paso serán manipulados por la clase *Utterance*.
  - La clase *Value* ordena los parámetros MIDI de entrada.
  - La clase *EmotionalConfig* contiene un método, *getMidi*, que guarda los valores que genera *RangeParams* en un objeto de tipo *Value*. Después, en función del valor de entrada *state* (CALMA, ALEGRIA, MIEDO...), devuelve la información MIDI asociada a una emoción (o *state*).
  - La clase *SetMidi* crea mensajes MIDI (objetos de la clase *Message* del paquete de Python Mido) con la información proporcionada por la clase *EmotionalConfig*.
  - La clase *Utterance* recibe los mensajes de *SetMidi* y los ordena en métodos que articulan la vocalización.

El comportamiento de la aplicación viene reflejado en el diagrama de secuencia UML de la Figura 4.4.

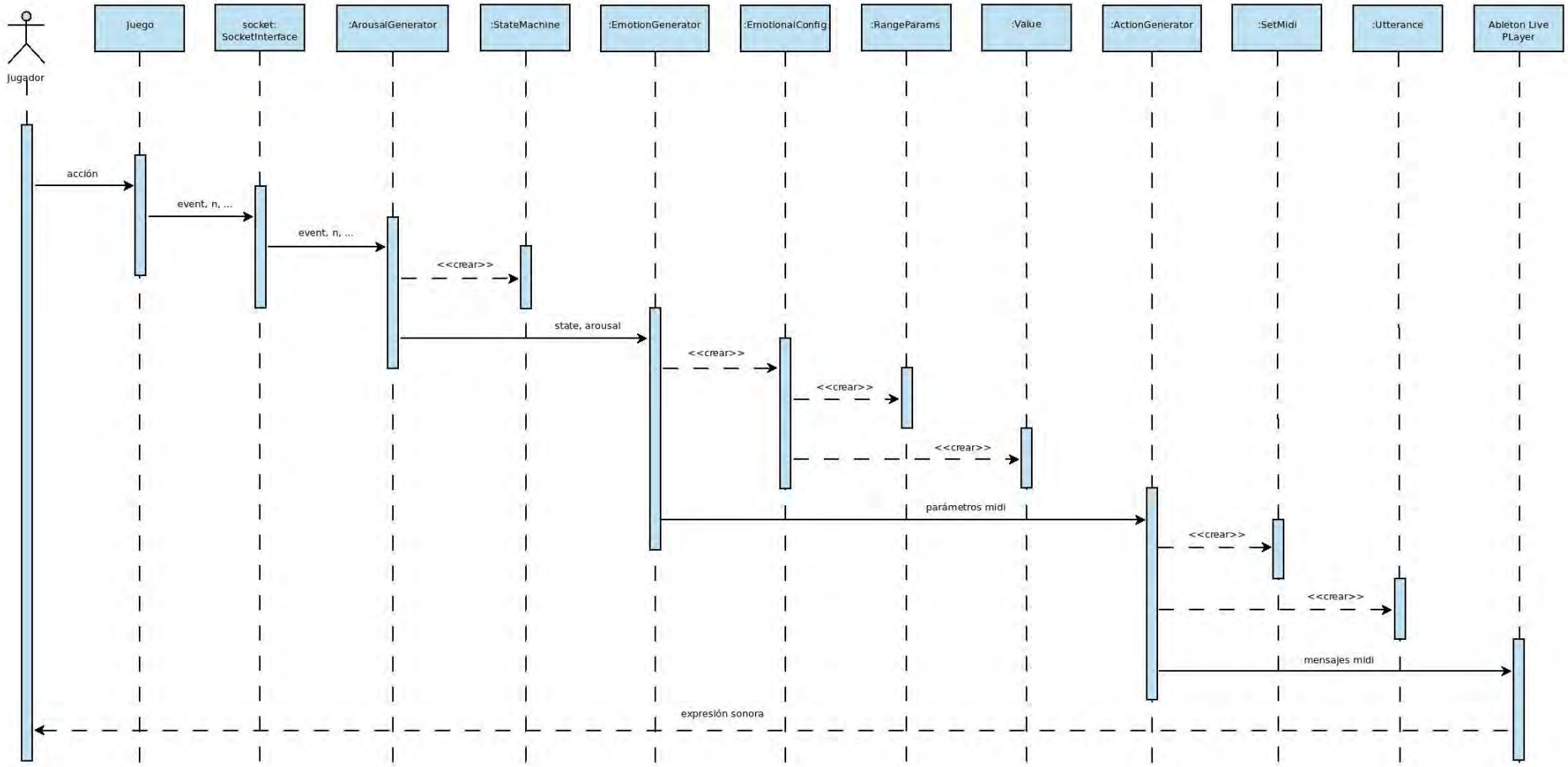


Figura 4.4: Diagrama de secuencia UML

4.2.2 Síntesis de expresiones en Ableton

Caracterización emocional

Antes pasar al siguiente apartado, es conveniente explicar la diferencia entre los dos tipos de parámetros sonoros que hemos configurado para caracterizar nuestras expresiones emocionales:

- *Parámetros estáticos*: aquellos que mantienen un valor constante durante todas las vocalizaciones.

- *Parámetros dinámicos*: aquellos que están mapeados a los *Controles Generales* y cuyos valores van a ser, durante la vocalización, modificados vía MIDI desde la aplicación en Python.

En este trabajo se distinguen dos fases en la caracterización. La primera fase se ha desarrollado en Ableton Live, y ha consistido en parametrizar la cobertura tímbrica y la entonación de cada expresión, por medio de parámetros estáticos. Así, en esta fase conseguimos el sonido distintivo que distinguirá una expresión emocional de otra.

La segunda fase de esta caracterización se ha desarrollado en Python, y se divide a su vez en dos subfases: a) la configuración de los parámetros diámicos y b) el control de estos parámetros -esta subfase englobará cualidades primarias como la inflexión y el ritmo, propias de cada expresión emocional-. Como vimos en el subapartado anterior, la clase *Utterance* realizaba la segunda fase de la caracterización.

Para un manejo más sencillo entre las expresiones sonoras, hemos creado dos grandes categorías emocionales: Placer y Displacer, y hemos asignado cada una a un track (elemento de Ableton que explicamos en el capítulo 3). En el track de Placer se encuentran las cadenas relativas a las expresiones de Alegría y Calma, y en el track de Displacer, una única cadena a partir del cual se generan las expresiones de Asco, Miedo y Tristeza (el porqué de esto último lo explicaremos en el siguiente apartado). Esta agrupación en categorías ha permitido trabajar de una forma más flexible e intuitiva, en términos de visualización y de mapeo MIDI.

Antes de comenzar la explicación hemos decidido, en aras de una buena comprensión de este capítulo, dividir esta sección en dos grandes apartados, siguiendo un modelo de explicación general - específico:

1. En el primer apartado, realizaremos una descripción comparativa de los parámetros sonoros estáticos de cada expresión sonora (cadena).

2. En el segundo apartado explicaremos la configuración general de cada rack y el uso de los controles generales para la parametrización dinámica en cada una de las cadenas.

Configuración de parámetros estáticos

En este apartado, como dijimos en la introducción de esta sección, compararemos los parámetros esenciales que hemos utilizado de manera estática para definir cada expresión sonora. Sólo comentaremos los parámetros más importantes para distinguir cada expresión, si bien hemos presentado todos los que se encuentran involucrados en el proceso.

Hay que tener en cuenta, además, que estos valores han sido ajustados empíricamente -a partir de la percepción subjetiva de los investigadores y de otros participantes- dentro de un rango teórico flexible.

Para que quede claro, mostramos nuevamente los parámetros que definen el comportamiento temporal de una envolvente, ya que en adelante mostraremos los valores que les hemos dado en Operator:

- *Ataque*: llamado *Attack* en Operator, determina el tiempo que el sonido parte de un valor mínimo a uno máximo en una característica concreta (frecuencia si es una envolvente de frecuencia, volumen si es una envolvente de volumen, etc).
- *Decaimiento*: llamado *Decay* en Operator, define el tiempo en el que el sonido alcanza el nivel de esta característica de la etapa de sostenido.
- *Sostenimiento*: llamado *Sustain* en Operator, especifica el valor de la característica al que permanecerá el sonido mientras se mantenga activo.
- *Relajación*: llamado *Release* en Operator, tiempo en el que la característica alcanza un valor nulo al desactivar el sonido.

Tras esto, en este apartado explicaremos el comportamiento de las cuatro envolventes que hemos utilizado para la síntesis de las expresiones: Dinámica o de volumen, del LFO, del Filtro de Frecuencia y de Tono.

1. *Envolvente Dinámica y Tipo de Onda*

   En la Tabla 4.3 podemos observar los valores de la envolvente de volumen de las tres cadenas creadas (Alegría, Calma y Displacer), así como el tipo de onda utilizada. Antes de explicar el por qué de estos valores, hay que resaltar que la expresión *Displacer* surge fruto de una modulación de dos ondas, como queda reflejado en la Tabla 4.3 y en la Figura 4.5.

Param[a] **/Cadena**	**Alegría**	**Calma**	**Displ (Modulada)**	**(Moduladora)**
Ataque (ms)	5.58	16.9	2.79	0
Decaimiento (s)	13.4	1.49	60	60
Sostenimiento (dB)	-4.1	0	0	0
Relajación (ms)	216	100	40	60000
Oscilador	Sw3	Sw3	Sw4	Triangular

Tabla 4.3: Parámetros de envolvente dinámica y tipo de onda

[a]Los valores de la variable Oscilador (*Oscillator* en Ableton) SwX, siendo X un número, corresponden a las formas de onda diente de sierra (acortamiento de Saw, en inglés).

Desde una perspectiva evolutiva hemos asumido que, en un contexto determinado, una expresión sonora cumplirá su función comunicativa en la medida en que adapte sus características sónicas al tipo de respuesta que se espera del interlocutor. Es decir, no sólo que el receptor entienda qué tipo de emoción se está mostrando, sino

que además, con ayuda de la situación estimular, pueda entender el tipo de acción que se le está demandando.

En el caso concreto de la envolvente de volumen, es de esperar que las expresiones que muestren urgencia o impliquen reforzamiento habrán de tener valores reducidos de *Ataque* y moderados-altos valores de *Decaimiento*. Éstos so los valores asociados a cada cadena:

Figura 4.5: Envolventes dinámicas Displacer

- *Displacer*: Respecto a las expresiones de Miedo o Asco, es preferible que el máximo de volumen se alcance cuanto antes (*Ataque* reducido), dado que en situación de peligro, unas centésimas de segundo pueden significar la diferencia entre sobrevivir o morir. Del mismo modo, hemos entendido que este máximo de volumen se mantenga un tiempo relativamente extenso (*Decaimiento* elevado) para servir de alerta a los que puedan escuchar.
- *Alegría*: el reforzamiento positivo de la conducta de matar cucarachas (causante de la alegría) aumenta en la medida en que la demora del reforzador (la expresión de alegría) se reduce (*Ataque* reducido). Por otra parte, valores moderadamente altos de *Decaimiento* permiten que la intensidad del reforzador sea mayor durante más tiempo. Ver Figura 4.6.
- *Calma*: no es necesario alcanzar el valor máximo de volumen con rapidez, ni mantener éste a lo largo del tiempo, por lo que los valores de *Ataque* y *Decaimiento* tenderán a ser menores y mayores, respectivamente, comparándolos con los casos anteriores.

Figura 4.6: Envolventes dinámicas Alegría/Calma

Nota: con *Decaimiento* no regulamos el tiempo de vocalización -esto lo hacemos con

Length-, sino que regulamos el descenso de volumen en función del tiempo desde su alcance máximo de pico.

En cuanto a la elección del tipo de onda (Figuras 4.7 y 4.8), decir que ésta se ha llevado a cabo en función de criterios de prueba-error, sin soporte teórico, excepto en el hecho de mantener el mismo tipo de onda (diente de sierra, con o sin modulación) para todas las expresiones, ya que éstas son ejecutadas con un timbre único propio de un sólo individuo.

Figura 4.7: Tipo de Oscilador Alegría/Calma

Figura 4.8: Tipo de Oscilador Displacer

2. *Envolvente del LFO*

El LFO (*Low-Frequency Oscillator*) es, en expresión vocal, uno de los elementos más potentes para generar la entonación y la apertura vocal que provoca la expresión de vocales. En nuestro estudio sólo ha sido utilizado para una cadena, la de *Displacer*, para generar, junto con la envolvente de tono (como veremos después), la entonación de las expresiones de *Asco*, *Miedo* y *Tristeza*. El valor de los parámetros de su envolvente pueden verse en la Tabla 4.4 y en la Figura 4.9.

Recordemos que un LFO creaba una variación de un parámetro del sonido en función de una forma de onda concreta. En este caso, nuestro LFO varía el tono de la vocalización en función de un seno, de modo que los valores de tono están comprendidos en un rango que sería la amplitud de onda, y la variación del tono respecto del tiempo viene determinada por un comportamiento senoidal.

La cantidad de este efecto depende del selector *Amount*, y la frecuencia, del ratio (*Rate*). La función de estos dos selectores hace referencia a la frecuencia de variación del parámetro y a la cantidad de esta variación, respectivamente. Veremos

Param/Cadena	Displacer
Ataque (ms)	0
Decaimiento (ms)	600
Sostenimiento (dB)	0
Release (s)	0.05
Wave shape	Sine

Tabla 4.4: Parámetros de la envolvente del LFO

cómo se utilizan en la sección correspondiente a la configuración de estos parámetros dinámicos en los nodos de ROS.

Los parámetros de la envolvente, en el caso del LFO, son conceptualmente más difíciles de comprender. El *Ataque* nulo indica que el comienzo de la variación de tono se produce al comienzo de la vocalización, y un *Decaimiento* tan corto (600 ms) mantiene el nivel de variación casi tan elevado como en el valor de pico. Esto provoca una vocalización extensa en cuanto a tono (con una amplia variación), ideal para mostrar gran expresividad en la amenaza o el lamento.

Figura 4.9: Envolvente de LFO Displacer

3. *Envolvente del Filtro de Frecuencia*

   En la síntesis de expresiones vocales, la envolvente del filtro de frecuencia (Tabla 4.5 y Figura 4.10) suele utilizarse para la apertura de las vocales, siguiendo la premisa de que las vocales abiertas, como la A o la E, contienen mayor número de frecuencias altas que las vocales cerradas, como la O o la U. Éstas son las características de cada cadena:

   - Tanto en *Alegría* como en *Calma*, la frecuencia de corte es relativamente alta, por lo que aunque dejará fuera las frecuencias más altas seguirá proporcionando un sonido abierto. Esta configuración es coherente con la idea de que las vocalizaciones relativas a emociones positivas pueden expresarse convenientemente con vocales abiertas. En el caso del *Displacer*, el selector *Frequency* está mapeado a uno de los controles generales, así que veremos su funcionamiento más adelante.
   - En cuanto a la envolvente del filtro de frecuencias propiamente dicha, decir que los valores de los parámetros correspondientes a la cadena *Displacer* hace que el

Param/Cadena	Alegría	Calma	Displacer
Ataque (ms)	0	0	57.5
Decaimiento (s)	10.4	25	1.65
Sostenimiento (%)	100	88	16
Relajación (s)	1	50	15.5
Freq (kHz)	7.02	7.02	Mapped

Tabla 4.5: Parámetros de la envolvente del filtro de frecuencia

valor máximo del filtro se alcance rápido (*Ataque* de 57,5 ms), pero que ese valor decaiga tras un segundo y medio y se mantenga ahí hasta su extinción. Esto ocasiona una apertura vocal al comienzo de la vocalización y después el cierre vocal, lo que permite crear la apariencia de un grito o llanto ahogado.

Figura 4.10: Envolvente del Filtro de Frecuencia Alegría/Calma/Displacer

4. *Envolvente de Tono*

Ésta es la herramienta de Operator más difícil de manejar, dada la complejidad operativa de los controles que la componen. Junto con el LFO, la envolvente de tono conforma la entonación de la vocalización. En la Tabla 4.6 y en la Figura 4.11 se muestran los valores de sus parámetros, que explicamos a continuación:

Param/Cadena	Alegría	Calma	Displacer
Ataque (ms)	0.35	-100 (Slope)	-100 (Slope)
Decaimiento (%)	Mapped	66 (Slope)	100 (Slope)
Sostenimiento (st)	+39	-48	+4
Relajación	11.8	100 (Slope)	100 (Slope)
Spread (%)	100	100	100
Transp (st)	Mapped	Mapped	+10

Tabla 4.6: Parámetros de la envolvente de tono

- *Alegría*: un *Ataque* positivo, unido a un efecto de *Envolvente de Tono (Pitch Envelope)* positivo y elevado, genera una entonación descendente, que, como vimos cuando explicamos los elementos paralingüísticos, transmite firmeza y confianza.
- *Displacer*: un *Ataque* negativo y elevado (-100 de pendiente) genera una entonación ascendente, lo que transmite indecisión o inseguridad.

Figura 4.11: Envolvente de Tono Alegría/Calma/Displacer

El resto de la configuración de parámetros no responde tanto a unas cualidades deter-

minantes en la diferenciación emocional de estas expresiones, como a características más periféricas y propias de el modo de expresar dichas emociones. El selector de apertura *Spread* extiende el sonido espacializándolo en estéreo, lo cual crea la sensación de sonido envolvente.

5. *Random en la cadena Calma*

   Esta vocalización se asemeja al partoleo de los periquitos cuando están cómodos. Analizando esta expresión vocal en varios individuos de esta especie, no descubrimos un patrón tonal, y tampoco uno rítmico, por lo que incorporamos a esta vocalización el efecto MIDI aleatorio *Random*, cuyo funcionamiento ya ha sido explicado. Ofrecemos en la Tabla 4.7 los valores de los selectores más relevantes.

Random	**Calma**
Chance (%)	55
Choices	3
Scale	3

Tabla 4.7: Efecto Random en la expresión de calma

El valor de alcance (*Chance*) implica que la probabilidad de que el valor MIDI de entrada sea sustituido por un valor aleatorio dentro de un rango obtenido del producto de *Elecciones (Choices)* y *Escalado (Scale)* es del 55%.

Configuración de parámetros dinámicos en los Controles Generales

1. *Rack Placer*

   Los controles generales del Instrument Rack de la Figura 4.12 controlan selectores de los efectos MIDI o de Operator a través del *Macro Map Mode* (que, como explicamos en el apartado dedicado a la explicación de Ableton live, permite la parametrización dinámica del sonido en tiempo real). Los selectores de los controles generales tienen un rango de valores de 0...127.

Figura 4.12: Controles del rack de Placer

- *Selector(Select)*: este control selecciona la cadena (*Alegría* o *Calma*) en la que se va a ejecutar la expresión emocional, actuando sobre los selectores de encendido de altavoces (*Speaker On*).
- *Transposición (Transp)*: este control actúa sobre el selector de Transpose de Operator en ambas cadenas.
- *Ratio (Rate)*: este control actúa sobre el selector de *Rate* del efecto MIDI *Arpeggiator* en ambas cadenas.
- *Longitud de nota (Length)*: este control actúa sobre el selector de *Time Length* del efecto MIDI *Note Length* en ambas cadenas.
- *Cantidad (Amount)*: este control actuaba sobre el selector de *Amount* de Operator. En la última versión de nuestro proyecto lo quitamos, ya que la variación del parámetro *Amount* ya no era necesaria.
- *Repeticiones (Repeats)*: este control actúa sobre el selector de *Repeats* del efecto MIDI *Arpeggiator* únicamente en la cadena *Alegría*.
- *Envolvente de Tono (Pitch. Env)*: este control actúa sobre el selector de *Amount* del Pitch Envelope de Operator en ambas cadenas.
- *Decaimiento*: este control actúa sobre el selector de *Decaimiento* del Pitch Envelope de Operator, en el caso de la cadena *Alegría*, y en el selector de *Time* en el caso de la cadena *Calma*.

Toda la información relativa al mapeo realizado con el *Macro Map Mode* puede verse en la Figura 4.13. A la derecha podemos ver los rangos de valores que pueden tomar cada uno de estos parámetros. Cogiendo el primero a modo de ejemplo (*Decaimiento* para Alegría), el mínimo valor (1 ms) corresponde al mínimo valor del selector *Decaimiento* explicado más arriba (0), y el máximo valor de este parámetro (60 s) correspondería al máximo valor del selector (127).

Figura 4.13: Macro Map Mode

Aunque en general esos valores máximo y mínimo los hemos mantenido por defecto, en otros parámetros hemos decidido adapatar nuestros propios valores, aprovechando la flexibilidad que nos ofrece Ableton (como se puede ver para el caso de *Rate* para Calma, o *Select* para Alegría).

Como podemos observar en la Figura 4.14, todos los controles generales controlan un selector en ambas cadenas (Alegría y Calma), exceptuando el caso de *Repeats*, que sólo lo hace en Alegría. Veremos ahora más detalladamente el control de los selectores en cada expresión.

Figura 4.14: Macro Mappings del rack de Placer

2. *Rack Displacer*

   En este rack el control (Figura 4.15) se realiza de la misma forma, si bien algunos controles generales controlan distintos selectores al anterior rack.

Figura 4.15: Controles del rack de Displacer

- *Select*: este control es otro ejemplo de control vestigial, ya que nos sirvió para controlar selectores de altavoces *Speaker On* cuando Miedo, Tristeza y Asco eran cadenas diferenciadas. Actualmente, después de que uniéramos esas tres cadenas en *Displacer* y decidiéramos diferenciar estas tres expresiones emocionales perviamente, en la parametrización de las emociones en un nodo ROS, todos los valores de *Select* apuntan al *Speaker On* de ésta.
- *Cantidad (Amount)*: este control actúa sobre dos selectores en Operator: *Time*

y *Amount* del LFO. Aquí tenemos un ejemplo de cómo mapear dos selectores en el mismo controlador general.

- *Ratio (Rate)*: este control actúa sobre el selector de *Rate* del LFO de Operator.
- *Longitud de nota (Length)*: este control actúa sobre el selector de *Time Length* del efecto MIDI *Note Length*.
- *Envolvente de Tono (Pitch. Env)*: este control actúa sobre el selector de *Amount* del Pitch Envelope de Operator.
- *Frecuencia de Filtro (Filter Freq.)*: este control actúa sobre el selector de *Filter Frecuency* de Operator.
- *Tono (Tone)*: este control actúa sobre el selector de *Tone* de Operator.
- *Macro 8*: este control no está mapeado a ningún selector y mantiene su nombre por defecto.

Podemos ver de nuevo el mapeo más claramente en la Figura 4.16.

Figura 4.16: Macro Mappings del rack de Displacer

Capítulo 5

Conclusiones

Como recordamos, los objetivos de este trabajo eran los siguentes:

- Desarrollar una aplicación software orientada a la interacción Humano-Robot, que capte estímulos del entorno, los procese y transforme en patrones de expresión sonora no verbal en tiempo real.
- Generar mediante síntesis de sonido un grupo de expresiones sonoras emocionales no verbales que serán emitidas durante el juego, en función del rendimiento del jugador.
- Implementar el sistema de interacción (juego + aplicación software + sonidos generados por síntesis) en Maggie.

Una vez finalizado el proyecto, comprobemos si hemos conseguido alcanzar lo que nos habíamos propuesto:

- ***Objetivo 1***: la aplicación software creada recibe información del juego "Kill the Croaches" (el número de cucarachas presentes y los eventos que suceden durante el juego, entre otros datos) y por medio de una serie de algoritmos los convierte en una expresión emocional específica, con un nivel de arousal acorde a la situación, que es emitida al jugador como feedback de su rendimiento en el juego.

  Gracias a ROS, un sistema software formado por una colección de herramientas y librerías, hemos podido generar una respuesta emocional en tiempo real, en función de las variables de salida del juego. Esto ha permitido que la interacción entre el jugador y Maggie sea posible.

- ***Objetivo 2***: por medio de las tecnologías que nos ha ofrecido el DAW Ableton Live -en concreto con el pluggin Operator-, hemos podido cumplir este objetivo, sintetizando a partir de formas de onda simples, envolventes, filtros y efectos MIDI, cada una de las expresiones emocionales que hemos necesitado para la realización de nuestro proyecto.

  Para la caracterización de cada expresión nos hemos basado en estudios precedentes comentados en este trabajo, en los que se estudiaban los efectos de las emociones primarias en los parámetros de la vocalización humana. En el caso de la calma, al no haber encontrado referencias claras sobre su parametrización sonora, hemos optado por basar su caracterización a partir de la observación de periquitos y su "parloteo" en estado de tranquilidad.

- ***Objetivo 3***: la implementación de la aplicación software, el juego y las expresiones emocionales generadas por síntesis fue realizada en Maggie con éxito, y se realizaron distintas pruebas para comprobar que todo funcionaba correctamente.

5.1 Desarrollos futuros

Una vez alcanzados los objetivos propuestos y dirigiendo la vista a investigaciones futuras, hemos reunido una serie de posibles mejoras o caminos nuevos por los que se podría seguir explorando:

- Programar distintos tipos de personalidad y modos de afrontamiento de las situaciones del juego, y permitir la opción de escoger el tipo de personalidad en un menú de opciones del juego..
- Tener en cuenta el fallo del usuario cuando intente destruir una cucaracha, afectando con ello a la respuesta emocional.
- Ingeniar un sencillo sistema de aprendizaje que tenga en cuenta, por un lado, la actividad/inactividad del jugador durante el juego, y por otro, el número de fallos y aciertos que ha tenido.
- Combinar la expresión emocional sonora con gestualización, para generar respuestas más complejas y expresivas.
- Realizar experimentos con usuarios para comprobar que las expresiones sonoras están realmente adecuadas a la emoción generada por la aplicación.
- Investigar si la expresión de emociones sirve como feedback para el usuario, y si puede llegar a modificar su rendimiento durante el juego.
- Incorporar una mayor variedad de vocalizaciones para expresar cada emoción.
- Incluir otras expresiones emocionales como enfado, sorpresa, y otro tipo de expresiones como aburrimiento, duda, silbidos, tarareos...

Bibliografía

[1] A. Öhman, *Psychophisiology of emotion: An evolutionary cognitive perspective.* Toronto: JAI Press: Advances in Psychophisiology, 1987.

[2] C. Breazeal, "Designing sociable robots.," *Massachusetts Institute of Technology*, 2002.

[3] A. Mehrabian, *Nonverbal Communication.* Illinois, Chicago: Aldine Atherton, 1972.

[4] C. Darwin, *La expresión de las emociones en el hombre y en los animales.* Madrid, Spain: Alianza Editorial, 1998.

[5] M. de los Ángeles Malfaz Vázquez, *Sistema de toma de decisiones basado en emociones y autoaprendizaje para agentes sociales autónomos.* PhD thesis, Escuela Politécnica Superior de la Universidad Carlos III de Madrid, 2007.

[6] B. R. Duffy, "Fundamental issues in affective intelligent social machines.," *The Open Artificial Intelligence Journal*, vol. 2, 2008.

[7] "Paro, therapeutic robot." `http://www.parorobots.com/`, 2016. [Online; acceso en 8 de Abril de 2016].

[8] "Aibo." `http://www.sony-aibo.com/`, 2016. [Online; acceso en 8 de Abril de 2016].

[9] "Necoro." `http://www.megadroid.com/Robots/necoro.htm`, 2016. [Online; acceso en 8 de Abril de 2016].

[10] "Jibo." `https://www.jibo.com/`, 2016. [Online; acceso en 9 de Abril de 2016].

[11] "Pepper: Softbank robotics." `https://www.ald.softbankrobotics.com/en/cool-robots/pepper`, 2016. [Online; acceso en 9 de Abril de 2016].

[12] K. Scherer and H. Walbott, "Evidence of universality and cultural variation of differential emotion response pattering.," *Journal of Personality and Social Psychology*, pp. pp. 310–328, 1994.

[13] K. Scherer, T. Johnstone, and G. Klasmeyer, "Vocal expression of emotion.," *Handbook of Affective Sciences*, pp. pp. 433–456, 2003.

[14] R. Read and T. Belpaeme, "How to use non-linguistic utterances to convey emotion in child-robot interaction.," *Centre for Robotics and Neural Systems*, 2012.

[15] R. Read, *A study of non-linguistic utterances for social human-robot interaction.* PhD thesis, Plymouth University, 2014.

[16] C. Gobl and A. Ní Chasaide, "The role of voice quality in communicating emotion, mood and attitude.," *Speech Communication*, 2003.

[17] J. Cacioppo and W. Gardner, *Emotion.* Anual Review of Psychology, 1999.

[18] R. Vallerand and C. Blanchard, *The study of emotion in sport and exercise.* Champaign: Human Kinetics, 2000.

[19] R. Lazarus, "Emotion and adaptation.," *Oxford University Press.*, 1991.

[20] A. Ortony, G. Clore, and A. Collins, "The cognitive structure of the emotions..," *Cambridge University Press*, 1988.

[21] P. Lang, *Fear reduction and fear behavior: Problems in treating a construct.* Washington D.C.: Research in psychotherapy, 1968.

[22] F. Poyatos, *La comunicación no verbal 1: Cultura, lenguaje y conversación.* Madrid: Biblioteca Lingüística y Filología Istmo, 1994.

[23] F. Poyatos, *Nonverbal Communication across Disciplines. Volume II: Paralanguage, kinesics, silence, personal and environmental interaction.* Amsterdam/Philadelphia: John Benjamins, 2002.

[24] S. Schachter and J. Singer, "Cognitive, social and psychological determinants of emotional state.," *Psychological Review*, 1962.

[25] P. Lang, M. Bradley, and B. Cuthbert, "International affective picture system (iaps): Technical manual and affective rating.," *The Center of Research in Psychophysiology, University of Florida*, 1999.

[26] M. Arnold, *Emotion and Personality. Vols. 1, 2.* Nueva York: Columbia University Press, 1960.

[27] P. Ekman, W. Friesen, and P. Ellsworth, *What are the similarities and differences in facial behavior across cultures?* Nueva York: Cambridge University Press, 1971.

[28] P. Ekman, "An argumet for basic emotions. cognition and emotion.," *University of California, San Francisco*, 1992.

[29] J. Dunn, *Emotional developement in early childhood: A social relatioship perspective.* Oxford: Oxford University Press, 2003.

[30] E. Gómez Gutiérrez, "Introducción a la síntesis de sonidos.," *Escola Superior de Musica de Catalunya*, 2009.

[31] R. A. García, "Automatic generation of sound synthesis techniques.," *Massachusetts Institute of Technology*, 2001.

[32] C. J. P., *The Sense of Hearing.* Lawrence Erlbaum Associates, 2005.

[33] T. Holmes, *Electronic and experimental music: technology, music, and culture.* Taylor & Francis, 2011.

[34] "Maggie: futuro, autonomía y diversión." `http://portal.uc3m.es/portal/page/portal/actualidad_cientifica/actualidad/reportajes/archivo_reportajes/Maggie_futuro_autonomia_diversion`, 2016. [Online; acceso en 9 de Abril de 2016].

[35] "Ros.org — powering the world's robots." `www.ros.org`, 2016. [Online; acceso en 24 de Marzo de 2016].

[36] "Python." `https://www.python.org/`, 2016. [Online; acceso en 24 de Marzo de 2016].

[37] I. Lütkebohle, "Bworld robot control software." `http://aiweb.techfak.uni-bielefeld.de/content/bworld-robot-control-software/`, 2008. [Online; accessed 19-July-2008].

[38] "Mido - midi objects for python." `https://mido.readthedocs.io/en/latest/`, 2016. [Online; acceso en 4 de Febrero de 2016].

[39] P. Ekman, *Basic Emotions*. John Wiley and Sons, 1999.